

Access DB# 68180**SEARCH REQUEST FORM**

Scientific and Technical Information Center

Requester's Full Name: \_\_\_\_\_ Examiner #: \_\_\_\_\_ Date: \_\_\_\_\_  
Art Unit: \_\_\_\_\_ Phone Number 30 \_\_\_\_\_ Serial Number: \_\_\_\_\_  
Mail Box and Bldg/Room Location: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

**If more than one search is submitted, please prioritize searches in order of need.**

\*\*\*\*\*  
Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_  
Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

*\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

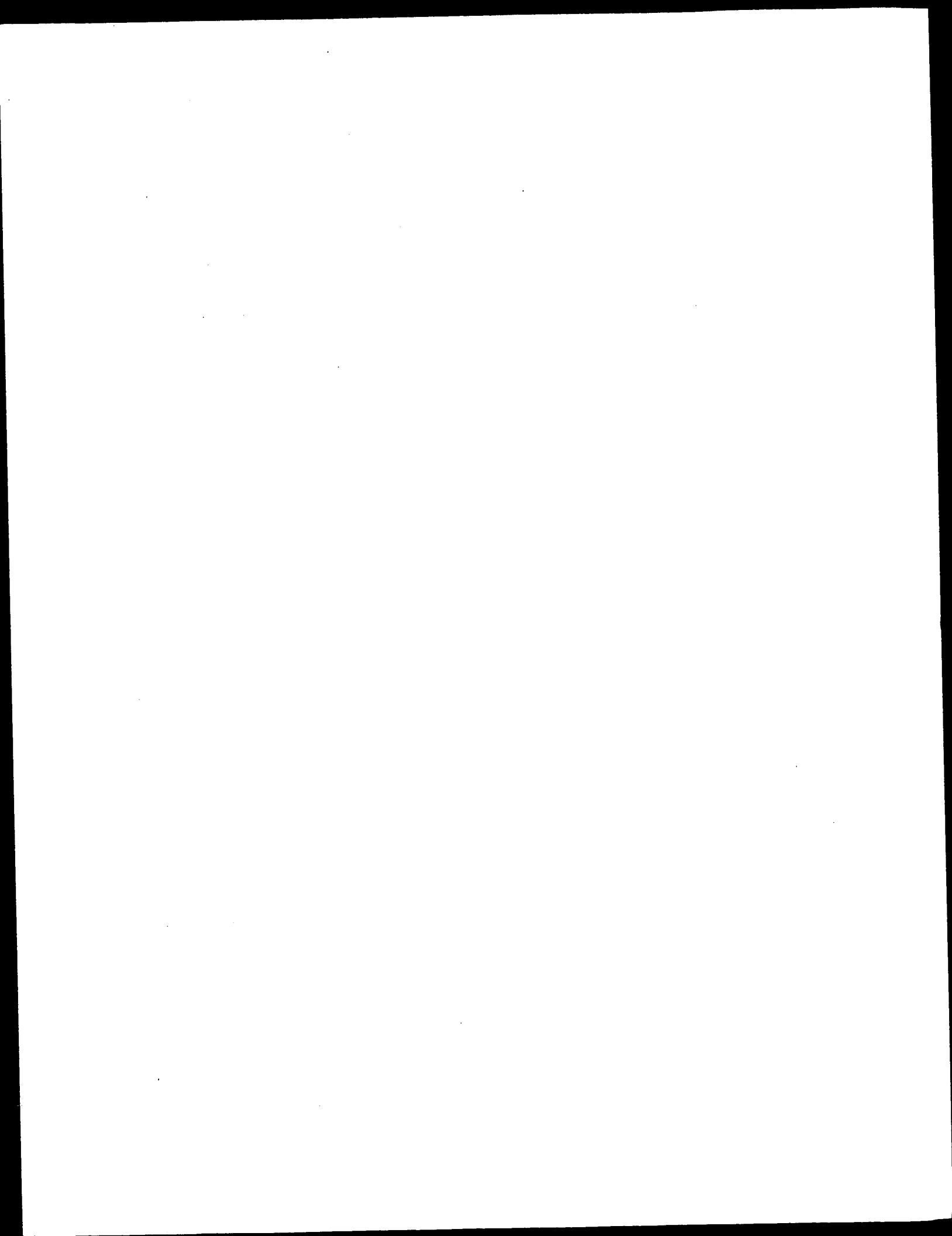
**BEST AVAILABLE COPY**

Jan Delaval  
Reference Librarian  
Biotechnology & Chemical Library  
CM1 1E07 - 703-308-4498  
jan.delaval@uspto.gov

**STAFF USE ONLY**

\*\*\*\*\*  
Searcher: Jan Type of Search \_\_\_\_\_ Vendors and cost where applicable  
Searcher Phone #: 2498 NA Sequence (#) 3 STN \_\_\_\_\_  
Searcher Location: \_\_\_\_\_ AA Sequence (#) 4 Dialog \_\_\_\_\_  
Date Searcher Picked Up: 6/6/02 Structure (#) \_\_\_\_\_ Questel/Orbit \_\_\_\_\_  
Date Completed: 6/6/02 Bibliographic \_\_\_\_\_ Dr. Link \_\_\_\_\_  
Searcher Prep & Review Time: \_\_\_\_\_ Litigation \_\_\_\_\_ Lexis/Nexis \_\_\_\_\_  
Clerical Prep Time: 15 Fulltext \_\_\_\_\_ Sequence Systems ☒  
Online Time: 10 Patent Family \_\_\_\_\_ WWW/Internet \_\_\_\_\_  
Other \_\_\_\_\_ Other (specify) \_\_\_\_\_

PTO-1590 (8-01)



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 06:09:37 ; Search time 2149.71 Seconds  
(without alignments)  
13054.080 Million cell updates/sec

Title: US-08-881-509-1

Perfect score: 1341

Sequence: 1 ATGAGGCAAGTGGCGAGAGT.....GCCGATGCCCTTCATTAATAAT 1341

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*

- 1: gb\_ba: \*
- 2: gb\_htg: \*
- 3: gb\_in: \*
- 4: gb\_om: \*
- 5: gb\_ov: \*
- 6: gb\_pat: \*
- 7: gb\_ph: \*
- 8: gb\_pl: \*
- 9: gb\_pr: \*
- 10: gb\_ro: \*
- 11: gb\_sts: \*
- 12: gb\_sy: \*
- 13: gb\_un: \*
- 14: gb\_vl: \*
- 15: em\_ba: \*
- 16: em\_fun: \*
- 17: em\_hum: \*
- 18: em\_in: \*
- 19: em\_mu: \*
- 20: em\_om: \*
- 21: em\_or: \*
- 22: em\_ov: \*
- 23: em\_pat: \*
- 24: em\_ph: \*
- 25: em\_pl: \*
- 26: em\_ro: \*
- 27: em\_sts: \*
- 28: em\_un: \*
- 29: em\_vi: \*
- 30: em\_htg\_hum: \*
- 31: em\_htg\_inv: \*
- 32: em\_htg\_other: \*
- 33: em\_htgo\_inv: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Score	Match Length	DB ID	Description
-----				

## ALIGNMENTS

RESULT 1

A93127

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

A93127

Sequence 1 from Patent EP0816496.

A93127

A93127.1

GI:6741516

unidentified.

unclassified.

Schudel,D.J.

1 (bases 1 to 1341)

T-cells specific for kidney carcinoma

Patent: EP 0816496-A 1 07-JAN-1998;

BOEHRINGER MANNHEIM GMBH (DE)

Location/Qualifiers

1..1341

/organism="unidentified"

/db\_xref="taxon:32644"

1..54

1..804

/note="unnamed protein product"

/codon\_start=1

linear

PAT 22-JAN-2000

A93127 Sequence 1  
X02592 Human mRNA  
M12959 Human T-cel  
BC020840 Homo sapi  
S72942 c6.1A...TCR  
S72931 Homo sapien  
M13052 Human T-cel  
AK026255 Homo sapi  
X01403 H.sapiens m  
X63455 H.sapiens m  
M15565 Human T-cel  
X05002 Human DNA f  
M14862 Human T-cel  
AL160314 Human chr  
AC023226 Homo sapi  
M12423 Human T-cel  
E00990 cDNA encod1  
AE000662 Homo sapi  
M94081 Human Tcr-C  
X02883 Human gene  
I14073 Sequence 8  
AX201131 Sequence  
AX267930 Sequence  
AF043177 Homo sapi  
AJ004871 Homo sapi  
Z49903 H.sapiens m  
AF043174 Homo sapi  
AX322052 Sequence  
L34703 Homo sapien  
AF043171 Homo sapi  
K02777 Human T-cel  
U40407 Synthetic c  
L02424 Homo sapien  
GI3536 SHGC-11082  
M55622 Sheep T cel  
D10394 Bovine T-ce  
M17663 Human T-cel  
AX010058 Sequence  
AX010083 Sequence  
AX010104 Sequence  
AX010157 Sequence  
AX010074 Sequence  
AX010093 Sequence  
AX010114 Sequence  
AX010173 Sequence

1	1341	100.0	1341	6	A93127
2	972.2	72.5	1508	9	HSTCRAR
3	972.2	72.5	1524	9	HUMTCAXB
4	958.2	71.5	1255	9	BC020840
5	950.8	70.9	1457	9	S72942
6	950.8	70.9	1647	9	S72931
7	835.4	62.3	874	9	HUMATGRA
8	691.4	51.6	1367	9	AK026255
9	583.2	50.9	1131	9	HSTCELA1
10	596.6	44.5	1064	9	HSTCARA
11	569.6	42.5	1080	9	HUMTCAYE
12	538.8	40.2	840	9	HSTCRAC4
13	538.8	40.2	840	9	HUMTCAC4
14	535.6	39.9	208953	9	CNS01RH3
15	534	39.8	175053	2	AC023226
16	514.8	38.4	1081	9	HUMTCAXE
17	514.8	38.4	1101	6	E00990
18	510	38.0	71153	9	HUAE000662
19	510	38.0	97630	9	HUMTCRADCV
20	498.2	37.2	5177	9	HSTCRAC
21	448	33.4	822	6	I14073
22	446.4	33.3	834	6	AX201131
23	446.4	33.3	834	6	AX267930
24	444.6	33.2	804	9	AF043177
25	441.6	32.9	828	9	HSJ004871
26	440.4	32.8	810	9	HSTCRDF8
27	440.2	32.8	813	9	AF043174
28	439.8	32.8	828	6	AX322052
29	439.2	32.8	840	9	HUMTCRAZ
30	438.2	32.7	858	9	AF043171
31	428	31.5	728	9	HUMTCAXA
32	423	31.3	426	9	HUMTCRA29Z
33	420.2	31.3	426	9	HUMTCRA29Z
34	393.4	29.3	400	11	GI3536
35	382.4	28.5	1410	4	SPTCRA
36	371.6	27.7	989	4	BOVTCRA21
37	335	25.0	397	9	HUMTCAYX
38	307.8	23.0	744	6	AX010058
39	307.8	23.0	744	6	AX010083
40	307.8	23.0	744	6	AX010104
41	307.8	23.0	744	6	AX010157
42	307.2	22.9	765	6	AX010074
43	307.2	22.9	765	6	AX010093
44	307.2	22.9	765	6	AX010114
45	307.2	22.9	765	6	AX010173

```

/protein_id="CAB69529.1"
/db_xref="GI:6741517"
/translation="MQQVARVFLTSLTSLAKTQPISMDSYEGEVNITCSHNNI
ATNDYITWYQQFSPQGRFLIOGYKTKTNEVASLIPADRKSSTLCLSPVLSLDTAV
YLCVLGSGARITFGGSGQTULPDIQNDPAVYLRDSKSKDSKVCILDTFDSQTNV
SQKSDSYAQLTFLVLDMSQFVSKNSAWKSDFAFANFNNSIIPEDTFPPSE
SSCDKLVKESFETDTNLNQLNSVIGFRILLKLVAGFNLLMLRLWSS"
55..801
/mat_peptide
/product="unnamed"
BASE COUNT 331 a 365 c 294 g 351 t
ORIGIN

```

Query Match	100.0%	Score 1341;	DB 6;	Length 1341;	
Best Local Similarity	100.0%;	Pred. No. 0;			
Matches 1341;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Qy	1	ATCAGCAAGTGGCGAGTGTATCGTGTCTCTGACCCCTGAGTACTTTGAGCCTTGCTAAG	60		
Db	1	ATGAGCAAGTGGCGAGTGTATCGTGTCTCTGACCCCTGAGTACTTTGAGCCTTGCTAAG	60		
Qy	61	ACCACCCAGCCCATCTCCATGGACTCATATGAAGGACAAGAAGTGAACATTAACCTGTAGC	120		
Db	61	ACCACCCAGCCCATCTCCATGGACTCATATGAAGGACAAGAAGTGAACATTAACCTGTAGC	120		
Qy	121	CACACAACATTTGCTACAAATGATTATATCAGCTGGTACCAACAGTTTCCAGCCCAAGGA	180		
Db	121	CACACAACATTTGCTACAAATGATTATATCAGCTGGTACCAACAGTTTCCAGCCCAAGGA	180		
Qy	181	CCAGGATTTATTATTCAGGGATACAGACAAAGTTTCAACAGAGTGGCTCCCTGTTT	240		
Db	181	CCAGGATTTATTATTCAGGGATACAGACAAAGTTTCAACAGAGTGGCTCCCTGTTT	240		
Qy	241	ATCCCTGCCGACAGAAGTCCAGCACTCTGAGCCTGCCCGGGTTTCCCTGAGCGACACT	300		
Db	241	ATCCCTGCCGACAGAAGTCCAGCACTCTGAGCCTGCCCGGGTTTCCCTGAGCGACACT	300		
Qy	301	GCTGTGTACTACTGCCTCGTGGGTGGTCTCGAAGGCAACTGACCTTTGGATCTGGGACA	360		
Db	301	GCTGTGTACTACTGCCTCGTGGGTGGTCTCGAAGGCAACTGACCTTTGGATCTGGGACA	360		
Qy	361	CAATTTGACTGTTTACCTGATATCCAGACCTGACCTGCCGCTGACCACTGAGAGAC	420		
Db	361	CAATTTGACTGTTTACCTGATATCCAGAACCTGACCTGCCGCTGACCACTGAGAGAC	420		
Qy	421	TCTAAATCCAGTGACAAAGTCTGTCTGCCATTATTCACCGATTTTGATTTCTCAACAATAATGTG	480		
Db	421	TCTAAATCCAGTGACAAAGTCTGTCTGCCATTATTCACCGATTTTGATTTCTCAACAATAATGTG	480		
Qy	481	TCACAAAGTAGGATTTCTGATGTGATATCAGACACAAACTGTGCTAGACATGAGGTCT	540		
Db	481	TCACAAAGTAGGATTTCTGATGTGATATCAGACACAAACTGTGCTAGACATGAGGTCT	540		
Qy	541	ATGACATTCGAAGACACAGTGTGTGGCTGGAGCAACAAATCTGACTTTGCATGTGCA	600		
Db	541	ATGACATTCGAAGACACAGTGTGTGGCTGGAGCAACAAATCTGACTTTGCATGTGCA	600		
Qy	601	AACGCCTTCAACACAGCATTTATTCAGAAAGACACTTCTTCCCGACGCCAGAAAGTTCC	660		
Db	601	AACGCCTTCAACACAGCATTTATTCAGAAAGACACTTCTTCCCGACGCCAGAAAGTTCC	660		
Qy	661	TGTGATGTCAGCTGTGTCGAGAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAAC	720		
Db	661	TGTGATGTCAGCTGTGTCGAGAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAAC	720		
Qy	721	CTGTCAAGTGATTTGGTTTCCGAATCCCTCCTGAAAGTGGCGGGTTTAAATCTGCTCATG	780		
Db	721	CTGTCAAGTGATTTGGTTTCCGAATCCCTCCTGAAAGTGGCGGGTTTAAATCTGCTCATG	780		
Qy	781	ACGCTGGCGGTGTGGTCCAGCTGAGATCTCGAGATTTGAACAGACGCTGTGCTCCCTCG	840		
Db	781	ACGCTGGCGGTGTGGTCCAGCTGAGATCTCGAGATTTGAACAGACGCTGTGCTCCCTCG	840		





Thu Jun 6 10:09:03 2002

Db	507	TAATCTTTGGATCAGGACAGACTCAGATCCGGCAATATCCAGAACCTTGACCCCTG	566
Qy	401	CGGTGTACAGCTGAGAGACTCTAAATCCAGTGCAGAGTCTGTGCGCTATTACCGATT	460
Db	567	CGGTGTACAGCTGAGAGACTCTAAATCCAGTGCAGAGTCTGTGCGCTATTACCGATT	626
Qy	461	TTGATCTCAACAATGTGTCAAAAGTAAAGTCTGTGATGTATATACAGACAAA	520
Db	627	TTGATCTCAACAATGTGTCAAAAGTAAAGTCTGTGATGTATATACAGACAAA	686
Qy	521	CTGTGTACAGATGAGTCTATGAGTCTCAAGAGCAACAGTGTGGCTGGAGCAACA	580
Db	687	CTGTGTACAGATGAGTCTATGAGTCTCAAGAGCAACAGTGTGGCTGGAGCAACA	746
Qy	581	AATCTGACCTTTCAGATGCAAAAGCTTCAACAACAGCATTTATCCAGAGAACCTTCT	640
Db	747	AATCTGACCTTTCAGATGCAAAAGCTTCAACAACAGCATTTATCCAGAGAACCTTCT	806
Qy	641	TCCCGACCCAGAAAGTCTCTGTGATGTCAGCTGGTGCAGAAAGCTTTGAAACAGATA	700
Db	807	TCCCGACCCAGAAAGTCTCTGTGATGTCAGCTGGTGCAGAAAGCTTTGAAACAGATA	866
Qy	701	CGAACCTTAACCTTCAAAAGCTCTGATGATTTGGTTCGGAATCTCTCTGAAAGTG	760
Db	867	CGAACCTTAACCTTCAAAAGCTCTGATGATTTGGTTCGGAATCTCTCTGAAAGTG	926
Qy	821	AGACAGCTGTCTCCCTGCTCTCTCTCTGATTTGGCTTCCGCTCTCTCTCTCAACA	880
Db	987	AGACAGCTGTCTCCCTGCTCTCTCTCTGATTTGGCTTCCGCTCTCTCTCTCAACA	1046
Qy	881	GAGGGAACCTCTCTACCCCAAGAGGTGAGCTGCTACACCTCTGTGCCCCCGGCT	940
Db	1047	GAGGGAACCTCTCTACCCCAAGAGGTGAGCTGCTACACCTCTGTGCCCCCGGCT	1106
Qy	941	AATGCCACCACTGGATCTACCCCAAGATTTATGATTAAGTGTGAAGAGCTGCCAACA	1000
Db	1107	AATGCCACCACTGGATCTACCCCAAGATTTATGATTAAGTGTGAAGAGCTGCCAACA	1166
Qy	1001	ACTGCTGCCACCCCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1060
Db	1167	ACTGCTGCCACCCCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1226
Qy	1061	AGGCAAGGCTGCTGACGCT	1120
Db	1227	AGGCAAGGCTGCTGACGCT	1286
Qy	1121	CTCTGCCATCCACAGATGATGATCTTCAGTGGGTTCTCTCTCTCTCTCTCTCTCTCTCT	1180
Db	1287	CTCTGCCATCCACAGATGATGATCTTCAGTGGGTTCTCTCTCTCTCTCTCTCTCTCTCT	1346
Qy	1181	GAATGCTGTGAGGGGTTTATTTTATAGTGTCTATAAGAAATACATAGTATCT	1240
Db	1347	GAATGCTGTGAGGGGTTTATTTTATAGTGTCTATAAGAAATACATAGTATCT	1406
Qy	1241	TCTTCTCAGAGCTGGGGGAAATATCTCATTTATCGAGCCCTCTGCTGTGCTCT	1300
Db	1407	TCTTCTCAGAGCTGGGGGAAATATCTCATTTATCGAGCCCTCTGCTGTGCTCT	1466
Qy	1301	GGCGGTGTGTATGCTCTGCTGCCGATGCCCTTCATTAAT	1341
Db	1467	GGCGGTGTGTATGCTCTGCTGCCGATGCCCTTCATTAAT	1507
RESULT	4		
LOCUS	BC020840	1255 bp	linear
DEFINITION	Homo sapiens, clone MGC:23964	IMAGE:4687209	mrna, complete cds.
ACCESSION	BC020840		

BC020840.1 GI:18088582  
MGC.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1. (bases 1 to 1255)  
Strausberg, R.  
Direct Submission  
Submitted (03-JAN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mdickson@stanford.edu](mailto:mdickson@stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 36 Row: k Column: 12  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis.  
Location/Qualifiers  
1. 1255  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:23964 IMAGE:4687209"  
/tissue\_type="testis, embryonal carcinoma"  
/clone\_lib="NIH-MGC\_61"  
/lab\_host="DH10B"  
/note="Vector: pDNR-LIB"  
66...674  
/codon\_start=1  
/product="Unknown (protein for MGC:23964)"  
/protein\_id="AAH20840.1"  
/db\_xref="GI:18088583"  
/translation="MORISLLHLFLWAGYMSAIELVPEHOTVPSIGVPTLRCSM  
KBAIGNIYINWYKTDQIDNPDPAYQVLRDSDKSKSVCLIFDQSTNVSQSKSD  
VYITDKTVIDMRSDPKSKSAVANSKDFACANAFNNSIIPDFFPSPSSCDVKL  
VERSFETDNLNPNLSVIGFRILLKLVAGFNLLMTLRWS"  
BASE COUNT 322 a 332 c 285 g 316 t  
ORIGIN  
Query Match 71.5%; Score 958.2; DB 9; Length 1255;  
Best Local Similarity 99.7%; Pred. No. 7e-296;  
Matches 960; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 379 GATATCCAGAACCCCTGACCCCTGCTGACGCTGACGACTCTAAATCCAGTGAACG 438  
Db 249 GATATCCAGAACCCCTGACCCCTGCTGACGCTGACGACTCTAAATCCAGTGAACG 308  
Qy 439 TCTGCTGCTTATTCACCATTTTGTATCTCAACAAATTTGTCTCAAAAGTAAAGTATCT 498  
Qy 309 TCTGCTGCTTATTCACCATTTTGTATCTCAACAAATTTGTCTCAAAAGTAAAGTATCT 368  
Db 499 GATGCTATATACAGACAAACTGTGTAGACATGAGGCTCTATGAGCTTCAAGAGCAAC 558  
Db 369 GATGCTATATACAGACAAACTGTGTAGACATGAGGCTCTATGAGCTTCAAGAGCAAC 428  
Qy 559 AGTCTGTGGCTGGACACAAATCTGACTTTGCTATGTGCAACGCCCTTCAACACAGC 618  
Db 429 AGTCTGTGGCTGGACACAAATCTGACTTTGCTATGTGCAACGCCCTTCAACACAGC 488



Thu Jun 6 10:09:03 2002

1097 TTCCCTCTCTCTCCCCAGAGACTGCTCCGCCCATCCAGATGATGATCTTCAGTGG 1156  
 1196 TTCCCTCTCTCTCCCCAGAGACTGCTCCGCCCATCCAGATGATGATCTTCAGTGG 1255  
 1157 TTCTCTGGGCTCTAGTCTCTGGAGATGTTGTGAGGGTTTATTTTAAATAGTGT 1216  
 1256 TTCTCTGGGCTCTAGTCTCTGGAGATGTTGTGAGGGTTTATTTTAAATAGTGT 1315  
 1217 TCATAAAGAAATACATAGTATCTTCTCTCAAGACGTGGGGGAAATATCTCATATC 1276  
 1316 TCATAAAGAAATACATAGTATCTTCTCTCAAGACGTGGGGGAAATATCTCATATC 1375  
 1277 GAGGCCCTCTATGCTGTGTG-TCTGGCGTGTGTATGCTCTGCGGATGCCCTTCAT 1335  
 1376 GAGGCCCTCTATGCTGTGTG-TCTGGCGTGTGTATGCTCTGCGGATGCCCTTCAT 1435  
 1336 TAAAT 1341  
 1436 TAAAT 1441

RESULT 6  
 S72931 1647 bp DNA linear PRI 08-MAY-2000  
 LOCUS Homo sapiens T-cell receptor alpha chain-c6.1A fusion protein  
 DEFINITION (c6.1A-TCRC) gene, partial cds.  
 ACCESSION S72931  
 VERSION S72931.1 GI:639612  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1647)  
 AUTHORS Thick,J., Mak,Y.F., Metcalfe,J., Beatty,D. and Taylor,A.M.  
 TITLE A gene on chromosome Xq28 associated with T-cell lymphocytic leukemia in two patients with ataxia telangiectasia  
 JOURNAL Leukemia 8 (4), 564-575 (1994)  
 MEDLINE 94202904  
 REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 154715] from the original journal article.  
 This sequence comes from Fig. 1.  
 Map location: t(X:14)(q28;q11).  
 FEATURES  
 source Location/Qualifiers  
 1..1647  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="2a"  
 /note="ataxia telangiectasia patient AT8BI"  
 <1..1093  
 /gene="c6.1A-TCRC"  
 <1..1093  
 /gene="c6.1A-TCRC"  
 /note="This sequence comes from Fig. 1; author's translation differs from conceptual translation"  
 /codon\_start=2  
 /product="T-cell receptor alpha chain-c6.1A fusion protein"  
 /protein\_id="AAB30469.2"  
 /db\_xref="GI:7717235"  
 /translation="DTRSDSKFAYTGMRTVAEKVDVIRVHSHVILRRDRKMD RVEIPEQLSASTAEARLAEITGRPMRVGVWVSHPHITVWPVSHVDVTRQAMQMD QGVGLIFSCFTEDKNTKTRGLVITCFQSQIAQKSSSLGPRDFWSSSHISIEGQK EEEVERIEIPHIIVPHVITIGKVCLESIAVELPKILCOEEDRYRTHSLHLSVTKI HNGSDIONDPAYQYLRDSKSDSKVCLDFDSDQTNVOSKSDSYITDKTVLDMRS MDPKNSAVAWNSKSDAFANFNNSIIPEDTFPPSPSSCDKLVKEKFEFTDNLNF ONLSVIGFRILLKLVAGFNLLMLRLWSS"  
 misc\_recomb 669^670  
 /gene="c6.1A-TCRC"  
 /note="translocation breakpoint"  
 BASE COUNT 427 a 422 c 373 g 425 t  
 ORIGIN

Query Match 70.9%; Score 950.8; DB 9; Length 1647;  
 Best Local Similarity 99.7%; Pred No. 1.8e-293;  
 Matches 963; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

377 CTGATATCCAGAACCTGACCCCTGCCCTGTACAGCTGAGAGACTCTAAATCCAGTGACA 436  
 666 CAGATATCCAGAACCTGACCCCTGCCCTGTACAGCTGAGAGACTCTAAATCCAGTGACA 725  
 437 AGTCTGCTGCTATTCACCGATTTTGAATTTCTCAAAACAAATGTGTCAAAAGTAAGGATT 496  
 726 AGTCTGCTGCTATTCACCGATTTTGAATTTCTCAAAACAAATGTGTCAAAAGTAAGGATT 785  
 497 CTGATGTGTATATCACAGACAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCA 556  
 786 CTGATGTGTATATCACAGACAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCA 845  
 557 ACAGTGTGTGGCTGGAGCAAAATCTGACTTTGCAATGTGCAAAACCCCTTCAACAACA 616  
 846 ACAGTGTGTGGCTGGAGCAAAATCTGACTTTGCAATGTGCAAAACCCCTTCAACAACA 905  
 617 GCATTTATCCAGAACCTGACCCCTGCCCTGTACAGCTGAGAGACTCTAAATCCAGTG 676  
 906 GCATTTATCCAGAACCTGACCCCTGCCCTGTACAGCTGAGAGACTCTAAATCCAGTG 965  
 677 TCGAGAAAGCTTTGAAACAGATACGAACTTAACTTCAAAACCTGTCAAGTATGGGT 736  
 966 TCGAGAAAGCTTTGAAACAGATACGAACTTAACTTCAAAACCTGTCAAGTATGGGT 1025  
 737 TCCGAATCTCTCTCTGAAAGTGGCGGGTTTAAATGTGTCTATGAGCGTGGCGGTGTGT 796  
 1026 TCCGAATCTCTCTCTGAAAGTGGCGGGTTTAAATGTGTCTATGAGCGTGGCGGTGTGT 1085  
 797 CCAGCTGAGATCTGCAAGATTTGAAGACAGCTGTCTCCCTGCTCTCTCTGCAAT 856  
 1086 CCAGCTGAGATCTGCAAGATTTGAAGACAGCTGTCTCCCTGCTCTCTCTGCAAT 1145  
 857 GCCCT 916  
 1146 GCCCT 1205  
 917 CTACCACT 976  
 1206 CTACCACT 1265  
 977 AGATTTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCTCTGTTCCCTTATTTGCTGCTT 1036  
 1266 AGATTTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCTCTGTTCCCTTATTTGCTGCTT 1325  
 1037 GTCACTGCTGACATTCACGGCAGAGGCAAGGCTGCTGACGCTCTCCCTGCTGTGCACA 1096  
 1326 GTCACTGCTGACATTCACGGCAGAGGCAAGGCTGCTGACGCTCTCCCTGCTGTGCACA 1385  
 1097 TTCCCT 1156  
 1386 TTCCCT 1445  
 1157 TTCTCTTGGGCTCTAGTCTCTGAGAAATGTGTGAGGGGTTTATTTTAAATAGTGT 1216  
 1446 TTCTCTTGGGCTCTAGTCTCTGAGAAATGTGTGAGGGGTTTATTTTAAATAGTGT 1505  
 1217 TCATAAAGAAATACATAGTATCTTCTCTCAAGACGTGGGGGAAATATCTCATATTC 1276  
 1506 TCATAAAGAAATACATAGTATCTTCTCTCAAGACGTGGGGGAAATATCTCATATTC 1565  
 1277 GAGGCCCTCTATGCTGTGTG-TCTGGCGTGTGTATGCTCTGCTGCGGATGCCCTTCAT 1335  
 1566 GAGGCCCTCTATGCTGTGTG-TCTGGCGTGTGTATGCTCTGCTGCGGATGCCCTTCAT 1625  
 1336 TAAAT 1341  
 1626 TAAAT 1631

```

RESULT 7
HUMATCRA
LOCUS      HUMATCRA      874 bp      DNA      linear      PRI 31-OCT-1994
DEFINITION Human T-cell antigen receptor alpha-chain mRNA, 3' end.
ACCESSION M13052
VERSION    M13052.1 GI:179135
KEYWORDS   T cell receptor
SOURCE     Human leukemic T-cell line Jurkat, cDNA to mRNA, clone pJ6-alpha-2.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Collins, M.K., Tanigawa, G., Kissnerghis, A.M., Ritter, M.,
            Price, K.M., Toney, S. and Owen, M.J.
            Regulation of T-cell receptor gene expression in human T-cell
            development
            Proc. Natl. Acad. Sci. U.S.A. 82 (13), 4503-4507 (1985)
FEATURES   Location/Qualifiers
            1..874
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /map="14q11.2"
            <1..>874
            /product="T-cell antigen receptor mRNA"
            1..323
            /gene="TCRA"
            <1..323
            /gene="TCRA"
            /note="T-cell antigen receptor alpha-chain"
            /codon_start=3
            /protein_id="AA51791.1"
            /db_xref="GI:179135"
            /db_xref="GDB:600-120-404"
            /translation="QSKDSYVITDKVLMRSDPKNSAVAWKNSDFACANAFNN
            STIPEDTFPSPSSCDVKLVKESFEDTNLFONLSVIGFRILLKLVAGFNLLMTLR
            LNSS"
BASE COUNT 206 a 238 c 195 g 235 t
ORIGIN      203 bp upstream of HindIII site; chromosome 14q11.2.
            Query Match      62.3%; Score 835.4; DB 9; Length 874;
            Best Local Similarity 99.7%; Pred. No. 1.8e-256;
            Matches 898; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
QY 482 CACAAAGTAAGGATCTCATGTGTATATACAGACAAACAACTGTCTAGACATGAGCTCTA 541
            |||||||
Db 1 CACAAAGTAAGGATCTCATGTGTATATACAGACAAACAACTGTCTAGACATGAGCTCTA 541
QY 542 TGGACTTCAAGAGCAACAGTCTGTGGCGTGGAGCAACAACTGTCTAGACATGAGCTCTA 601
            |||||||
Db 61 TGGACTTCAAGAGCAACAGTCTGTGGCGTGGAGCAACAACTGTCTAGACATGAGCTCTA 601
QY 602 AGCCCTTCAACACAGCATATTCAGAGACACCTTCTTCCCGCCAGCAAGTCTCT 661
            |||||||
Db 121 AGCCCTTCAACACAGCATATTCAGAGACACCTTCTTCCCGCCAGCAAGTCTCT 661
QY 662 GTGATGTCAAGCTGGTGGAGAAAGCTTTGAAACACATACGACCTTAACTTTCAAAACC 721
            |||||||
Db 181 GTGATGTCAAGCTGGTGGAGAAAGCTTTGAAACACATACGACCTTAACTTTCAAAACC 721
QY 722 TGTCAAGTATGGTTCGGAATCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 781
            |||||||
Db 241 TGTCAAGTATGGTTCGGAATCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 781
QY 782 CGCTGGCGGTGTGGTCCAGCTGAGATCTCAGATTTGAAGACAGCTGTCTCTCTCTCTCG 841
            |||||||
Db 301 CGCTGGCGGTGTGGTCCAGCTGAGATCTCAGATTTGAAGACAGCTGTCTCTCTCTCTCG 841
QY 842 TCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 901
            |||||||
Db 361 TCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 920

```

```

QY 902 AGGAGGTCAAAGCTGTCTACCACTCTCTGTGCCCCCGGCAATGCCACCACTGATCCTA 961
            |||||||
Db 421 AGGAGGTCAAAGCTGTCTACCACTCTCTGTGCCCCCGGCAATGCCACCACTGATCCTA 961
QY 962 CCCGAATTTATGATTAAAGATTGCTGAAGAGCTGCCAAACACACACCCCTCTCTCT 1021
            |||||||
Db 481 CCCGAATTTATGATTAAAGATTGCTGAAGAGCTGCCAAACACACACCCCTCTCTCT 1021
QY 1022 CCCTTATGCTGCTGTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1081
            |||||||
Db 541 CCCTTATGCTGCTGTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1081
QY 1082 CCCTGCTGTGCACATTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
            |||||||
Db 601 CCCTGCTGTGCACATTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
QY 1141 ATGATCTTCAGTGGTGGTCTCTTTGGGCTCTAGGCTCTGAGAAATGTTGTGAGGGGTTTAT 1200
            |||||||
Db 661 ATGATCTTCAGTGGTGGTCTCTTTGGGCTCTAGGCTCT -GAGAAATGTTGTGAGGGGTTTAT 1200
QY 1201 TTTTAAATAGTGTTCATTAAGAAATACATAGTATCTTCTTCTCAAGACGTGGGGG 1260
            |||||||
Db 720 TTTTAAATAGTGTTCATTAAGAAATACATAGTATCTTCTTCTCAAGACGTGGGGG 1260
QY 1261 AAATATCTCATATTCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
            |||||||
Db 780 AAATATCTCATATTCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
QY 1321 TCCGATGCTTCATTAATAAT 1341
            |||||||
Db 840 TCCGATGCTTCATTAATAAT 860

RESULT 8
AK026255 Homo sapiens CDNA: FLJ22602 fis, clone HSI04504, highly similar to
LOCUS      HSTCARA Homo sapiens mRNA for T-cell antigen receptor alpha-chain.
DEFINITION AK026255
ACCESSION  AK026255.1 GI:10439050
VERSION     AK026255
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens human small intestine CDNA to mRNA, clone_lib:HSI
            clone:HSI04504.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Watanabe, K., Kumagai, A., Itakura, S., Yamazaki, M., Tashiro, H.,
            Ota, T., Suzuki, Y., Iobayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
            Nakamura, Y., Iobayashi, M., Nishi, T., and Sugano, S.
            NEDO human cDNA sequencing project
            Unpublished (2000)
            2 (bases 1 to 1367)
            Sugano, S., Suzuki, Y., Ota, T., Iobayashi, M., Nishi, T., Isogai, T.,
            Shibahara, T., Tanaka, T., and Nakamura, Y.
            Direct Submission
            Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumlo
            Sugano, Institute of Medical Science, University of Tokyo,
            Laboratory of Genome Structure Analysis, Human Genome Center;
            Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
            (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
            NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing; Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing; Department of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES     Location/Qualifiers
            1..1367
            /organism="Homo sapiens"

```

	1038	895	1093	955	1093	1015	1093	1075	1093	1135	1127	1195	1187	1255	1247	1315	1307	
b																		
b	CCCTCGCTCTCTCTCTGCAATGCCCCCTCTCTCTCCCTCTCTCAACAGAGGGAAC	ACCCCAAGGAGGTGAAGCTGTACCAACCTCTGTGCGCCCCCGGCAATGCCACCACTG	-----	GATCTACCCGAAATTTATGATTAAGATTCGTGAAGAGCTGCCAACAACACTGCTGCCACCC	-----	CTCTGTTCCTCTTATTTGCTGTGCTTGTCACCTGCCTGACATTCACGGCAGAGCAAGCTGCTG	-----	CAGCTCCCTCGCTGTGCACATTCCTCTCGCTCTCTCGCTCCCGCAGAGCTGCTCGCCATCCCA	-----	CAGCTCCCTCGCTGTGCACATTCCTCTCGCTCTCTCGCTCCCGCAGAGCTGCTCGCCATCCCA	-----	CAGATGATGATCTTCAGTGGGTCTCTCTGGGCTCTAGTCTCTGAGAGATTTGTGTAGGG	-----	CAGATGATGATCTTCAGTGGGTCTCTCTGGGCTCTAGTCTCTGAGAGATTTGTGTAGGG	-----	GGGGGAAATATCTCATTTATCAGTGGCTCTAGTCTCTGAGAGATTTGTGTAGGG	-----	TCCTGCTGCCGATCCCTTCATTAATAAT
b	1092	1092	1092	1092	1092	1074	1092	1134	1092	1134	1126	1194	1186	1254	1246	1314	1306	
b																		
b	1092	1092	1092	1092	1092	1074	1092	1134	1092	1134	1126	1194	1186	1254	1246	1314	1306	
b																		
b	1092	1092	1092	1092	1092	1074	1092	1134	1092	1134	1126	1194	1186	1254	1246	1314	1306	
b																		
b	1092	1092	1092	1092	1092	1074	1092	1134	1092	1134	1126	1194	1186	1254	1246	1314	1306	
b																		
b	1092	1092	1092	1092	1092	1074	1092	1134	1092	1134	1126	1194	1186	1254	1246	1314	1306	
b																		
b	1092	1092	1092	1092	1092	1074	1092	1134	1092	1134	1126	1194	1186	1254	1246	1314	1306	
b																		
b	1092	1092	1092	1092	1092	1074	1092	1134	1092	1134	1126	1194	1186	1254	1246	1314	1306	
b																		
b	1092	1092	1092	1092	1092	1074	1092	1134	1092	1134	1126	1194	1186	1254	1246	1314	1306	
b																		
b	1092	1092	1092	1092	1092	1074	1092	1134	1092	1134	1126	1194	1186	1254	1246	1314	1306	
b																		
b	1092	1092	1092	1092	1092	1074	1092	1134	1092	1134	1126	1194	1186	1254	1246	1314	1306	
b																		
b	1092	1092	1092	1092	1092	1074	1092	1134	1092	1134	1126	1194	1186	1254	1246	1314	1306	
b																		
b	1092	1092	1092	1092	1092	1074	1092	1134	1092	1134	1126	1194	1186	1254	1246	1314		

QY	926	CTGTGCCCCCGCGCAAT-GCCACCAACTGGATCCTACCGGATTTATGATTAGAT-TG	983
Db	947	CTGTGCCCCCGCGCAATGCCACCAACTGGATCCTACCGGATTTATGATTAGATGTG	1006
QY	984	CTGAAGAGCTGCCAACAACACTGTGCACACCCCTCTGTTCCTTATTGCTGTGTGTCAC	1043
Db	1007	CTGAAGAGCTGCCAACAACACTGTGCACACCCCTCTGTTCCTTATTGCTGTGTGTCAC	1066
QY	1044	CTGTGACATTACCGGCAGAGCAAGCTGTGCAGACCTCCCTGCTGTGCACATCCCTC	1103
Db	1067	CTGTGACATTACCGGCAGAGCAAGCTGTGCAGACCTCCCTGCTGTGCACATCCCTC	1125
QY	1104	CTGCTC 1109	
Db	1126	CTGCTC 1131	
RESULT	10		
HSTCARA			
LOCUS		HSTCARA	
DEFINITION		H.sapiens mRNA for T-cell antigen receptor alpha-chain.	
ACCESSION		X63455	
VERSION		X63455.1	
KEYWORDS		influenza haemagglutinin peptide specific; T-cell antigen receptor (alpha chain); V alpha 1.2.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		1 (bases 1 to 1064)	
AUTHORS		Hewitt,C.R.A.	
TITLE		Submitted (09-DEC-1991) C.R.A. Hewitt, St. Mary's Hospital Medical School, Dept.Of Immunology, Norfolk Place, Paddington, London	
JOURNAL		W21PG, UK	
REFERENCE		2 (bases 1 to 1064)	
AUTHORS		Hewitt,C.R., Lamb,J.R., Hayball,J., Hill,M., Owen,M.J. and O'Hehir,R.E.	
TITLE		Major histocompatibility complex independent clonal T cell energy by direct interaction of Staphylococcus aureus enterotoxin B with the T cell antigen receptor	
JOURNAL		J. Exp. Med. 175 (6), 1493-1499 (1992)	
MEDLINE		92268797	
FEATURES			
source		Location/Qualifiers	
		1..1064	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="UB alpha 14/4"	
		/cell_line="HA1.7"	
		/cell_type="T lymphocyte"	
		1..89	
		90..917	
5'UTR		/codon_start=1	
CDS		/product="T cell antigen receptor alpha chain"	
		/protein_id="CAA43053.1"	
		/db_xref="GI:36731"	
		/translation="MLLLVPLEVIFTLGGTGAQSVTQLGSHVSVSEGALVLLRCNY SSSVPLYFWYVOYVNOGLOLLKVTSAITLVKINGEAFEFKSETSFHLTKPSAHM SDAAEYFCVASVSPFENKLTGTGTRLTIIPTNTPDPVAYQLRDSKSDKSVCLFT DEDSQTNVQSQSDSVYITDKVLDMRDFKSNVAWSNKSDFACANAFNNSITPE DTFFFSPSSCDVKLVKSPETDNTLNQNLVSIGFRILLKLVAGFNLLMTLRWSS"	
sig_peptide		90..149	
mat_peptide		150..914	
misc_feature		/product="T cell antigen receptor alpha chain"	
		150..427	
misc_feature		/note="V region"	
		428..488	
misc_feature		/note="J /N region"	
		489..917	
		/note="C region"	
3'UTR		918..1064	
BASE COUNT		258 a 307 c 240 g 259 t	
ORIGIN			



Thu Jun 6 10:09:03 2002

FEATURES		Location/Qualifiers	
source		1..1080	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/map="14q11.2"	
		1..1080	
gene		/gene="TCRA"	
sig_peptide		109..189	
		/gene="TCRA"	
		/note="T-cell receptor signal peptide"	
		109..951	
CDS		/gene="TCRA"	
		/note="T-cell receptor precursor"	
		/codon_start=1	
		/protein_id="AAA60627.1"	
		/db_xref="GI:338766"	
		/db_xref="GDB:G00-120-404"	
		/translation="MAMLLGASVLIWLOPDMVNSQKNDQVKNQSPSPSLSVQEGRI SILNCDYNSMFDYFLWYKYPAGETFLISISIKDKNEDGRFTVFLNKAHLSLH IVPQSDSAYIFCAKAGAGTASKLFTGGTGLQVLTLDIQNDPDPAYKQDRSKSDKS VCLFTDFSDQTNVSDKSDVYIITDKTYLDMRSMDFKSNVAVNSKNSDFACANFNN SILPEDTFPPSPSSCDVKLVKSFETDINLNFONLSVIGFRILLKLVAGFNLLMTLR LWSS"	
mat_peptide		190..948	
		/gene="TCRA"	
		/product="T-cell receptor"	
		468..469	
misc_recomb		/gene="TCRA"	
		/organism="Homo sapiens"	
		524..525	
misc_recomb		/gene="TCRA"	
		/organism="Homo sapiens"	
BASE COUNT		303 a 274 c 239 g 264 t	
ORIGIN		Chromosome 14q11.2.	
		Query Match 42.5%; Score 569.6; DB 9; Length 1080;	
		Best Local Similarity 84.2%; Pred. No. 4.3e-171;	
		Matches 670; Conservative 0; Mismatches 114; Indels 12; Gaps 2;	
QY	141	TGATTATATACAGTGGTACCAACAGTTTCCAGCCAGGACACACGATTTATTATTCAAGG	200
DB	276	TGATTATTTCCTATGGTACAAAATAACCTGCTGAAGGTCTTACATTCCTGATATCTAT	335
QY	201	ATACAGACAAAGTTACAAACGAAGTGGCTCCCTGTTTATCCCTGCGACGACAAAGTC	260
DB	336	AAAGTTCCATTAGGATAAAATAAGATGAAGATTCACTGTCTTCTTAAACAAAGTGC	395
QY	261	CAGCACTCTGACCTGCGCCCGGTTTCCTCCAGCCAGTGTGCTGTCTGTACTACTG	314
DB	396	CAAGCACTCTCTCTGACATTTGTCCTCCAGCTGAGAGACTCTGAGTGTACTCTG	455
QY	315	-----CCTCGTGGTGGTTCGCAAGCAACTGACCTTTGGATCTGGGACACAATGAC	368
DB	456	TGCAGCAAAAGGGCGCGGCTGCGAGTAACTCACCTTTGGGACTGGAACAGACTTCA	515
QY	369	TGTTTACCTGATATCCAGAACCTTGACCTGCGCTGTACCCAGCTGAGAGACTTAATC	428
DB	516	GGTCAAGCTGATATCCAGAACCTTGACCTGCGCTGTACCCAGCTGAGAGACTTCAATC	575
QY	429	CAGTGAAGTCTGTCTGCTGCTATTCACCGATTTTGAATCTCAACAAATGTGTCAAAAG	488
DB	576	CAGTGAAGTCTGTCTGCTGCTATTCACCGATTTTGAATCTCAACAAATGTGTCAAAAG	635
QY	489	TAAGTATTCATGTAT	548
DB	636	TAAGTATTCATGTAT	695
QY	549	CAAGTGAAGTCTGTCTGCTGCTGCGCTGGAGCAACAAATCTGACTTTGATGTGCAAGCCTT	608
DB	696	CAAGTGAAGTCTGTCTGCTGCTGCGCTGGAGCAACAAATCTGACTTTGATGTGCAAGCCTT	755
QY	609	CAACAAAGTAT	668

RESULT 11  
 HUMTCAYE 1080 bp mRNA linear PRI 13-JAN-1995  
 LOCUS Human T-cell receptor rearranged alpha-chain V-region (V-D-J) mRNA,  
 DEFINITION complete cds.  
 ACCESSION M15565  
 VERSION 1 GI:338765  
 KEYWORDS T-cell receptor antigen; cell surface glycoprotein.  
 SOURCE Human cytotoxic T-lymphocyte, cDNA to mRNA, clone L1711-alpha.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1080)  
 Leiden, J.M., Fraser, J.D. and Strominger, J.L.  
 The complete primary structure of the T-cell receptor genes from an  
 alloreactive cytotoxic human T-lymphocyte clone  
 Immunogenetics 24 (1), 17-23 (1986)  
 JOURNAL 86276770  
 MEDLINE



```

Db 756 CACACAGCATTATTCCAGAGACACCTTCTCCAGCCAGAAAGTTCTGTGATGT 815
QY 669 CAAGCTGGTGCAGAAAAGCTTTGAACAGATAGACACCTAAATTTCAAAACCTGTCA 728
Db 816 CAAGCTGGTGCAGAAAAGCTTTGAACAGATAGACACCTAAATTTCAAAACCTGTCA 875
QY 729 GATTGGGTTCCGAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 788
Db 876 GATTGGGTTCCGAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 935
QY 789 GCTGTGGTCCAGCTGAGATCTGCAAGATTTGAACAGACCTGTGCTCCCTGCTCTTCC 848
Db 936 GCTGTGGTCCAGCTGAGATCTGCAAGATTTGAACAGACCTGTGCTCCCTGCTCTTCC 995
QY 849 TCTGTATGCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTCC 908
Db 996 TCTGTATGCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTCC 1055
QY 909 GAAAGCTGCTACCACC 924
Db 1056 GAAAGCTGCTACCACC 1071

RESULT 12
HSTRAC4
LOCUS
DEFINITION Human DNA for T-cell receptor constant region alpha-chain exon 4.
ACCESSION X05002 M14861
VERSION X05002.1 GI:36917
KEYWORDS constant region; T-cell receptor; T-cell receptor alpha.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 840)
AUTHORS Baer, R., Lefranc, M.P., Minowada, J., Forster, A., Stinson, M.A. and
Rabbitts, T.H.
TITLE Organization of the T-cell receptor alpha-chain gene and
rearrangement in human T-cell leukaemias
JOURNAL Mol. Biol. Med. 3 (3), 265-277 (1986)
MEDLINE 86284199
COMMENT See also X02883 for complete TCR C(alpha) gene with some
differences in exon 4 sequence.
FEATURES
source
location/Qualifiers
1..840
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="B-cell DNA library"
/clone="lamdaalpa3"
<1..34
/notes="Intron III"
35..36
/number=4
misc_feature 553..558
/notes="put. polyA signal"
BASE COUNT 175 a 246 c 186 g 233 t
ORIGIN

Query Match 40.2%; Score 538.8; DB 9; Length 840;
Best Local Similarity 97.8%; Pred. No. 3.1e-161;
Matches 546; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 784 CTGGCGCTGTGGTCCAGCTGAGATCTGCAAGATTTGAACAGACCTGTGCTCCCTGCTC 843
Db 13 CTTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 72
QY 844 CTTCTCTGTGATTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 903
Db 73 CTTCTCTGTGATTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 132
QY 904 GAGTGAAAGCTGCTACCACCTCTGTGCCCCCCCCCGGCAATGCCAAGTGGATCTCTACC 963

```

```

Db 133 GAGGTGAAAGCTGCTACCACCTCTGTGCCCCCGGCAATGCCAAGTGGATCTCTTACC 192
QY 964 CGAATTTATGATTGAAGTGTGAAGAGCTGCCAAACACTGCTGCCACCCCTCTGTTC 1023
Db 193 CGAATTTATGATTGAAGTGTGAAGAGCTGCCAAACACTGCTGCCACCCCTCTGTTC 252
QY 1024 CTTATTGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1083
Db 253 CTTATTGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 312
QY 1084 CTGGCTGTGCACATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1143
Db 313 CTGGCTGTGCACATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 372
QY 1144 GATCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1203
Db 373 GATCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 432
QY 1204 TTTTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1263
Db 433 TTTTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 492
QY 1264 TTATCTCATTCAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1323
Db 493 TTATCTCATTCAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 552
QY 1324 CGATGCTTCAATTAAT 1341
Db 553 CGATGCTTCAATTAAT 570

HUMTCAC4
LOCUS
DEFINITION Human T-cell receptor germline alpha-chain C-region gene, exon 4,
clone lambda-D-alpha-3.
ACCESSION M14861
VERSION M14861.1 GI:338712
KEYWORDS C-region; T-cell receptor; germline.
SEGMENT 4 of 4
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 840)
AUTHORS Baer, R., Lefranc, M.P., Minowada, J., Forster, A., Stinson, M.A. and
Rabbitts, T.H.
TITLE Organization of the T-cell receptor alpha-chain gene and
rearrangement in human T-cell leukaemias
JOURNAL Mol. Biol. Med. 3 (3), 265-277 (1986)
MEDLINE 86284199
COMMENT Clean copy of sequence [1] kindly provided by M.-P. Lefranc
(01-AUG-1986).
FEATURES
location/Qualifiers
1..840
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="14q11.2"
/germline
join(M14858.1:1..300,M14859.1:1..120,M14860.1:1..240,
1..558)
/genes="TCRA"
<1..34
/genes="TCRA"
/notes="TCA Intron C"
35..568
/partial
/genes="TCRA"
/note="G00-120-404"
BASE COUNT 175 a 246 c 186 g 233 t
ORIGIN About 0.5 kb after segment 3.

```

Thu Jun 6 10:09:03 2002

```

Query Match      40.2%; Score 538.8; DB 9; Length 840;
Best Local Similarity 97.8%; Pred. No. 3.1e-161;
Matches 546; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 784 CTGGGGCTGGTCCAGCTCAGATGTCAGAGATTGTAAGACAGCCCTGTGCTCCCTCGCTC 843
Db 13 CTTCTGTTCTTCCCTCATTCAGATCTGCAAGATTGTAAGACAGCCCTGTGCTCCCTCGCTC 72
Qy 844 CTTCTGTCATGTCCTCTCTTCCCTCTCCAAACAGAGGAGGACCTCTCTACCCCCCAAG 903
Db 73 CTTCTGTCATGTCCTCTCTTCCCTCTCCAAACAGAGGAGGACCTCTCTACCCCCCAAG 132
Qy 904 GAGGTGAAGCTGTACCACTCTGTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 963
Db 133 GAGGTGAAGCTGTACCACTCTGTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 192
Qy 964 CGAATTTATGATTGAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCTCTCTCC 1023
Db 193 CGAATTTATGATTGAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCTCTCTCC 252
Qy 1024 CTTATGCTGCTTGTCACTGCTGACATTCACGCGAGAGGCAAGGCTGCTGCAGCCCTCC 1083
Db 253 CTTATGCTGCTTGTCACTGCTGACATTCACGCGAGAGGCAAGGCTGCTGCAGCCCTCC 312
Qy 1084 CTGGCTGTGCATATCCCTCTGCTGCCAGAGACTGCTCCGCCATCCACACAGATGATG 1143
Db 313 CTGGCTGTGCATATCCCTCTGCTGCCAGAGACTGCTCCGCCATCCACACAGATGATG 372
Qy 1144 GATCTTCACTGGTCTCTTGGGCTCTAGGCTGCGAGAAATGTTGTGAGGGGTTTATTTT 1203
Db 373 GATCTTCACTGGTCTCTTGGGCTCTAGGCTGCGAGAAATGTTGTGAGGGGTTTATTTT 432
Qy 1204 TTTTATATGTTTATAAAGAAATACATGATGATTTCTTCTCAAGACGTGGGGGAAA 1263
Db 433 TTTTATATGTTTATAAAGAAATACATGATGATTTCTTCTCAAGACGTGGGGGAAA 492
Qy 1264 TTATCTATATCGAGCCCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1323
Db 493 TTATCTATATCGAGCCCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 552
Qy 1324 CGATGCTTCAATAAAT 1341
Db 553 CGATGCTTCAATAAAT 570

RESULT 14
CNS01RH3 208953 bp DNA linear PRI 07-DEC-2001
LOCUS Human chromosome 14 DNA sequence BAC C-2555K7 of library CalTech-D
DEFINITION from chromosome 14 of Homo sapiens (human), complete sequence.
ACCESSION AL160314
VERSION AL160314.7 GI:17426585
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 208953)
Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,
Brottier, P., Catolico, L., Barbe, V., Pelletier, F., Artiguenave, F.,
Levy, M., Eckenberg, R., Bruls, T., deBerardinis, V., Cruaud, C.,
Gyapay, G., Saurin, W. and Weissbach, J.
Sequencing of the human chromosome 14
Unpublished
2 (bases 1 to 208953)
Genoscope.
Direct Submission
Submitted (05-DEC-2001) Genoscope - Centre National de Sequencage :
Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
On Dec 8, 2001 this sequence version replaced gi:16945087.
----- Genome Center

```

```

Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr
-----
The following BAC sequence is oriented from the T7 to the SP6 end.
Downstream BAC (overlapping the SP6 end) : R-137H15 (AC-AL135998)
----- Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 10.49x in Q20 bases; sum-of-contigs
-----
Overall quality chart :
Range : bases
0 :
1 - 9 :
10 - 19 : 7
20 - 29 : 23
30 - 39 : 108
40 - 49 : 590
50 - 59 : 1914
60 - 69 : 3072
70 - 79 : 9115
80 - 89 : 37378
90 - 99 : 156746
-----
Percentage of bases with a quality value >= 40 : 99 %
-----
FEATURES
source
Location/Qualifiers
1..208953
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone_lib="CalTech-D"
11348..11777
/note="matching EMBL:M14862
RHdb:RH11152
dbSTS:STS328
Identified using the e-PCR software (G. Schuler)"
20600..20768
/note="matching EMBL:G13536
RHdb:RH13835
dbSTS:STS1067
Identified using the e-PCR software (G. Schuler)"
33864..34045
/note="matching EMBL:D12193
RHdb:RH96131
RHdb:RH1200
dbSTS:STS15139
Identified using the e-PCR software (G. Schuler)"
64772..64875
/note="matching EMBL:R42714
RHdb:RH53928
dbSTS:STS21457
Identified using the e-PCR software (G. Schuler)"
81032..81213
/note="matching EMBL:T50419
RHdb:RH91439
dbSTS:STS64574
Identified using the e-PCR software (G. Schuler)"
81063..81213
/note="matching EMBL:G14776
RHdb:RH8029
dbSTS:STS2224
Identified using the e-PCR software (G. Schuler)"
60111 a 46575 c 45879 g 56388 t
BASE COUNT
ORIGIN

```

```

Query Match      39.9%; Score 535.6; DB 9; Length 208953;
Best Local Similarity 97.5%; Pred. No. 1.1e-159;
Matches 544; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```



Thu Jun 6 10:09:03 2002

\* 91850 112964: contig of 21115 bp in length  
 \* 112965 113064: gap of 100 bp  
 \* 113065 134547: contig of 21483 bp in length  
 \* 134548 134647: gap of 100 bp  
 \* 134648 175053: contig of 40406 bp in length.

FEATURES  
 source  
 1. .175053  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="14"  
 /map="14"  
 /clone="RP11-576P2"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 1. .1065

misc\_feature  
 /note="assembly\_fragment"  
 1166. .2413  
 /note="assembly\_fragment"  
 2514. .4839  
 /note="assembly\_fragment"  
 4940. .6965  
 /note="assembly\_fragment"  
 7086. .8841  
 /note="assembly\_fragment"  
 8942. .11475  
 /note="assembly\_fragment"  
 11576. .14560  
 /note="assembly\_fragment"  
 14661. .19338  
 /note="assembly\_fragment"  
 19439. .22444  
 /note="assembly\_fragment"  
 22545. .27238  
 /note="assembly\_fragment"  
 clone\_end:T7  
 vector\_side:left  
 27339. .33032  
 /note="assembly\_fragment"  
 33133. .38008  
 /note="assembly\_fragment"  
 38109. .43595  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:right  
 43696. .51575  
 /note="assembly\_fragment"  
 51676. .59667  
 /note="assembly\_fragment"  
 59768. .69201  
 /note="assembly\_fragment"  
 69302. .79384  
 /note="assembly\_fragment"  
 79485. .91749  
 /note="assembly\_fragment"  
 91850. .112964  
 /note="assembly\_fragment"  
 113065. .134547  
 /note="assembly\_fragment"  
 134648. .175053  
 /note="assembly\_fragment"  
 2012 others

BASE COUNT 49542 a 38101 c 38584 g 46814 t  
 ORIGIN

Query Match 39.8%; Score 534; DB 2; Length 175053;  
 Best Local Similarity 97.3%; Pred. No. 3.4e-159;  
 Matches 543; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
 QY 784 CTGGCGGTGTGGTCCAGTCTGCAAGATTGTAAAGACAGCCCTGCTCCCTCGCTC 843  
 Db 75682 CTTCTGTCTCTCCATCTCCAGATCTGCAAGATTGTAAAGACAGCCCTGCTCCCTCGCTC 75741  
 QY 844 CTTCTGTCTCTCCCTCTCTCCCTCTCCAAACAGAGGGAACCTCTCTACCCCAAG 903

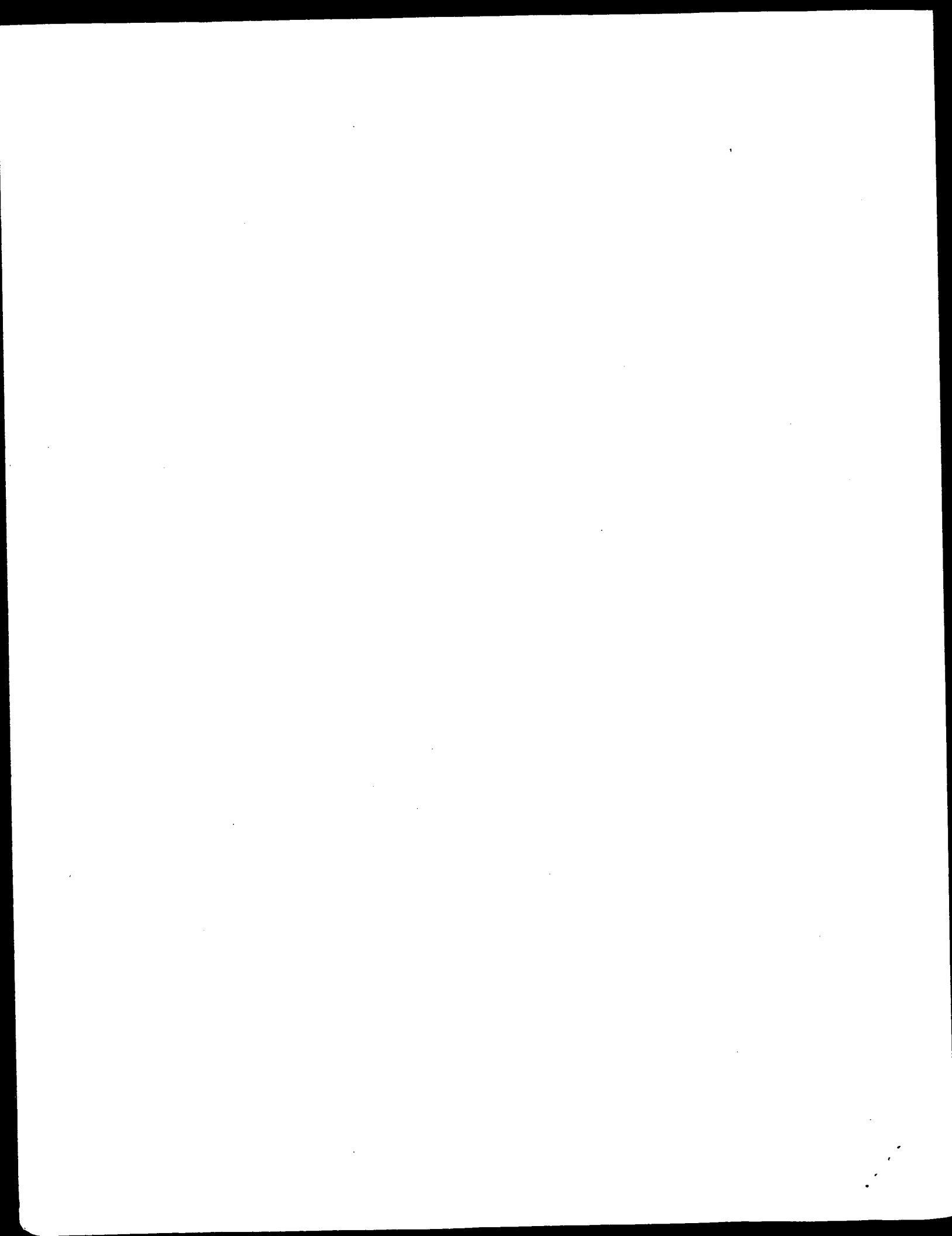
Db 75742 CTTCTCTGCAATTGCCCTCTCTCCCTCTCCAAACAGAGGGAACCTCTCTACCCCAAG 75801  
 QY 904 GAGGTGAAAGCTGTACCACTCTGTGCCCCCGGCAATGCCCAACTGATGCTCTACC 963  
 Db 75802 GAGGTGAAAGCTGTACCACTCTGTGCCCCCGGCAATGCCCAACTGATGCTCTACC 75861  
 QY 964 CGAATTATGATTAAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCTCTGTTC 1023  
 Db 75862 CGAATTATGATTAAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCTCTGTTC 75921  
 QY 1024 CTTATTGCTCTTTGTCACTGCTGCAATTCACGGCAGAGGCAAGGCTGCTGCAGCTCCC 1083  
 Db 75922 CTTATTGCTCTTTGTCACTGCTGCAATTCACGGCAGAGGCAAGGCTGCTGCAGCTCCC 75981  
 QY 1084 CTGGCTGTGCACATTCCTCTCTCCAGAGACTGCTCCGCCATCCACAGATGATG 1143  
 Db 75982 CTGGCTGTGCACATTCCTCTCTCCAGAGACTGCTCCGCCATCCACAGATGATG 76041  
 QY 1144 GATCTTCAGTGGTTCCTTTGGGCTCTAGGTCTCCGGGAGATGTTGTGAGGGGTTTATTT 1203  
 Db 76042 GATCTTCAGTGGTTCCTTTGGGCTCTAGGTCTCCGGGAGATGTTGTGAGGGGTTTATTT 76101  
 QY 1204 TTTTAAATAGTGTCAATAAAGAAATACATAGTATTTCTTCTCAAGACGTGGGGGAAA 1263  
 Db 76102 TTTTAAATAGTGTCAATAAAGAAATACATAGTATTTCTTCTCAAGACGTGGGGGAAA 76161  
 QY 1264 TTATCTCATTTATCAGGCGCTGCTATGCTGTGTGCGGCTGTTGTATGCTCTGCTGC 1323  
 Db 76162 TTATCTCATTTATCAGGCGCTGCTATGCTGTGTGCGGCTGTTGTATGCTCTGCTGC 76221  
 QY 1324 CGATGCCCTTCATTTAAAT 1341  
 Db 76222 CGATGCCCTTCATTTAAAT 76239

Search completed: June 6, 2002, 07:39:08  
 Job time: 5371 sec

Thu Jun 6 10:09:03 2002

us-08-881-509-1.rge

Page 15



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 06:09:37 : Search time 2649.09 Seconds  
(without alignments)  
6832.315 Million cell updates/sec

Title: US-08-881-509-1  
Perfect score: 1341  
Sequence: 1 ATGAGCGAAGTCCGAGAGT.....GCCGATGCTTCATTAAAT 1341

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estlin.\*  
4: em\_estnu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hc.\*  
12: gb\_gss.\*  
13: em\_gss\_hum.\*  
14: em\_gss\_inv.\*  
15: em\_gss\_pln.\*  
16: em\_gss\_vrt.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	972.6	72.5	1468	11	BC020270 Homo sapi
2	768.6	57.3	1012	10	BM479883 AGENCOURT
3	730.6	54.5	834	10	BI911817 603065260
4	722.8	53.9	869	10	BI764928 603047318
5	704.6	52.5	929	10	BG684892 602636627
6	694.4	51.8	822	9	AI816828 wj34f05.x
7	688.4	51.3	967	10	BI911009 603068729
8	651.8	48.6	717	10	BI772543 603054462
9	632.8	47.2	1054	10	BM457321 AGENCOURT
10	620.8	46.3	913	10	BI524586 603052009
11	611.4	45.6	973	10	BG719964 602691317
12	603	45.0	871	10	BG756578 602713728
13	593.2	44.2	638	9	AW514300 h474f09.x
14	584	43.5	749	9	AI351789 qt09c03.x
15	578	43.1	621	9	AW410870 f08h05.y
16	577	43.0	617	9	AW512331 xx73c07.x
17	570.8	42.6	610	9	AW976912 EST389021

C 18	563.6	42.0	618	9	AW026377	w12e05.x
C 19	556.4	41.5	619	9	AI660195	w61g07.x
C 20	555.6	41.4	721	9	AI357224	qx17h02.x
C 21	551.8	41.1	692	9	AI471656	tl199a09.x
C 22	547.6	40.8	762	10	BG758732	BG758732 602712974
C 23	534.4	39.9	589	9	AW263612	xq46h02.x
C 24	534	39.8	582	9	AI813335	wj33e12.x
C 25	530.8	39.6	595	10	BI907181	BI907181 603065776
C 26	516.4	38.5	887	10	BM457146	BM457146 AGENCOURT
C 27	510.8	38.1	558	9	AI741470	AI741470 wglb02.x
C 28	500	37.3	908	10	BI771427	BI771427 603059463
C 29	471.4	35.2	551	9	AW130991	AW130991 xe8c04.x
C 30	471.2	35.1	599	10	BE254521	BE254521 601115131
C 31	470.8	35.1	527	9	AA854890	AA854890 aj78e02.x
C 32	464.8	34.7	485	9	AI094938	AI094938 qal8d09.x
C 33	464.6	34.6	756	10	BG393458	BG393458 602411822
C 34	462.8	34.5	518	9	AW591225	AW591225 xoi2901.x
C 35	462.6	34.5	565	9	AA400994	AA400994 zu50f04.s
C 36	462	34.5	505	9	AA916904	AA916904 OH85a07.s
C 37	460.8	34.4	511	9	AI831179	AI831179 wj81h09.x
C 38	460.6	34.3	491	9	AW450946	AW450946 UI-H-B13-
C 39	456.8	34.1	554	10	BG257692	BG257692 602377574
C 40	446.2	33.3	495	9	AI091979	AI091979 oz96f11.x
C 41	445.2	33.2	488	9	AA284727	AA284727 zt24d05.s
C 42	440.4	32.8	800	10	BG826688	BG826688 602750823
C 43	440.4	32.8	896	10	BI833474	BI833474 603088138
C 44	438.4	32.7	569	10	BE258557	BE258557 601108335
C 45	437	32.6	458	9	AI718656	AI718656 as94g02.x

## ALIGNMENTS

RESULT 1

BC020270

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC020270 Homo sapiens, clone IMAGE:4764815, mRNA.  
BC020270.1 GI:17939628  
HTC  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1468)  
Strausberg, R.  
Direct Submission  
Submitted (19-DEC-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Louis Staudt  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
BC Cancer Agency, Vancouver, BC, Canada  
info@bcgc.bc.ca  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
Leticia Hsiao, Martin Krzyzinski, Reta Kutsche, Oliver Lee, Soo  
Ness, Pawan Pandoh, Carrie Mathewson, Candice McLeavy, Steven  
Schein, Duane Smailus, Anna-Liisa Prabhu, Parvaneh Seedi, Jacqueline  
Michael Thorne, Miranada Tsai, Michael Smith, Lorraine Spence, Jeff Stott,  
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 40 Row: h Column: 7  
This clone was selected for full length sequencing because it





QY	499	GATGTGTATATACAGACAAAACCTGCTGACACATGAGGTCTATGGACATTCAGAGCAAC	558
Db	263	GATGTGTATATACAGACAAAACCTGCTGACACATGAGGTCTATGGACATTCAGAGCAAC	322
QY	559	AGTGCCTGGCCCTGGAGCACAATCTGACTTTGCATGTGCAACGCCCTTCAACACAGC	618
Db	323	AGTGCCTGGCCCTGGAGCACAATCTGACTTTGCATGTGCAACGCCCTTCAACACAGC	382
QY	619	ATTATPCCAGAGACACCCCTCTCCCCAGCCGCAAGAGTTCCTGTGATGTCAAGCTGGTC	678
Db	383	ATTATPCCAGAGACACCCCTCTCCCCAGCCGCAAGAGTTCCTGTGATGTCAAGCTGGTC	442
QY	679	GAGAAAAGCTTTGAACACAGATACGAACTAAACTTTCAAAAGCTGTCACTGATTCGGGTTC	738
Db	443	GAGAAAAGCTTTGAACACAGATACGAACTAAACTTTCAAAAGCTGTCACTGATTCGGGTTC	502
QY	739	CGAATCTCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATCACGCTCGGGCTGTGGPCC	798
Db	503	CGAATCTCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATCACGCTCGGGCTGTGGPCC	562
QY	799	AGCTGAGATCTGCAAGATTCGTAAGACAGCCTGTGCTCCCTGCTCTCTCTCTCATTTGC	858
Db	563	AGCTGAGATCTGCAAGATTCGTAAGACAGCCTGTGCTCCCTGCTCTCTCTCTCATTTGC	622
QY	859	CCCTCTTCTCCCTCTCCAAACAGAGGAACTCTCTACCCCCAGGAGGTGAAGCTGCT	918
Db	623	CCCTCTTCTCCCTCTCCAAACAGAGGAACTCTCTACCCCCAGGAGGTGAAGCTGCT	682
QY	919	ACCACCTCTGTGCCCCCGCCGCAATGCCACACCTGGATCTACCCGAAATTTATGATTA	978
Db	683	ACCACCTCTGTGCCCCCGCCGCAATGCCACACCTGGATCTACCCGAAATTTATGATTA	742
QY	979	GATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCTCTGTTCCTTATTCGTGTGT	1038
Db	743	GATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCTCTGTTCCTTATTCGTGTGT	802
QY	1039	CACCTGCTGACATTCACGGCAGAGCAGGCTGCTGCAGCCTCCCTGGCTGTGCACATT	1098
Db	803	CACCTGCTGACA - TCACGGCAGAGCAGGCTGCTGCAGCCTCCCTGGCTGTGCACATT	861
QY	1099	CCCTCTGCTCCCGAGACATGCCCTCCGCCATCCACAGATGATGATCTTCAGTGGGT	1158
Db	862	TGCTCTGGCT - CCCAGAGACTGCCCTGGCCATCCAACAGATGATGATCTTCAGTGGGT	920
QY	1159	CTCTTTGGGCTCTAGTGCTCTGGAGAACTCTCTGAGGGGTTATTTTTTAAATAGTGTTC	1218
Db	921	CTCTTTGGGCTCTAGTGCTCTGGAGAACTCTCTGAGGGGTTATTTTTTAAATAGTGTTC	978
QY	1219	ATAAAGAATAACATAGTATTTCTTCTCTCAAGA	1251
Db	979	CNTAAGAAATTTACTAGTATTTCTTCTCTCAAGA	1011

```

RESULT      3
BI911817
LOCUS
DEFINITION  834 bp mRNA linear EST 16-OCT-2001
             603063260FL NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5214375 5',
             sequence.
ACCESSION   BI911817
VERSION     BI911817.1 GI:16175689
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 834)
             NTH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-remail.nih.gov
             Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.

```

	933	CCCCGGCAATGCCACCAACTGGATCTTACCAGAAATTATGATTAGATTGCTGAAGAC	992
DQ			
	636	CCCCGGCAATGCCACCAACTGGATCTTACCAGAAATTATGATTAGATTGCTGAAGAC	695
Db			
	993	TGCGAACAAGTGCTGCCACCCCTCTGTTCCTTTATTTGCTGCTTGCTCAGCTGCACATT	1052
QY			
	696	TGCGAACAAGTGCTGCCACCCCTCTGTTCCTTTATTTGCTGCTTGCTCAGCTGCACATT	755
Db			
	1053	CACGCGAGAGGAGGCTGCTGCAGCTCCCTCGCTGTGCACATTCCTCTGCTCCCTCCC	1112
QY			
	756	CACGCGAGAGGAGGCTGCTGCAGCTCCCTCGCTGTGCACATT-CTTCCTGCTCCCTCCC	814
Db			
	1113	AGAGACTGC	1121
QY			
	815	AGAGACTGC	823
Db			
	RESULT	4	
	BI764928	BI764928	869 bp mRNA linear EST 25-SEP-2003
	LOCUS	603047318F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5187413	5'
	DEFINITION	mRNA sequence.	
	ACCESSION	BI764928	
	VERSION	BI764928.1 GI:15756506	
	KEYWORDS	EST.	
	SOURCE	human.	
	ORGANISM	Homo sapiens	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
	REFERENCE	1 (bases 1 to 869)	
	AUTHORS	NIH-MGC http://mgc.nci.nih.gov/	
	TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)	
	JOURNAL	Unpublished (1999)	
	COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs@email.nih.gov Tissue procurement: Life Technologies, Inc. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA sequencing by: Invitrogen, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM11468 row: d column: 06 High quality sequence stop: 763. Location/Qualifiers	
	FEATURES	source	
		1..869	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:5187413"	
		/lab_lib="NIH_MGC_116"	
		/lab_host="PHI0B"	
		/note="Organ: pooled colon, kidney, stomach; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH_MGC Library."	
	BASE COUNT	228 a 239 c 188 g 214 t	
	ORIGIN		
	Query Match	53.9%; Score 722.8; DB 10; Length 869;	
	Best Local Similarity	95.2%; Pred. No. 2.7e-186;	
	Matches	811; Conservative 0; Mismatches 32; Indels 9; Gaps 6	
	QY	422 CTAAATCCAGTGACAGTCTGTCCTCATTCACCGATTTTGAATTCCTCAACAAATGTT	481
	Db	1 CTAAATCCAGTGACAGTCTGTCCTCATTCACCGATTTTGAATTCCTCAACAAATGTT	60



Thu Jun 6 10:09:08 2002

KEYWORDS SOURCE ORGANISM EST. human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 967) NIH-MGC http://mhc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Tissue Procurement: Life technologies, Inc. CDNA Library Preparation: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Cloning Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

FEATURES source 1. 967 /organism="Homo sapiens" /db\_xref="taxon:9606" /clone="IMAGE:5217870" /clone\_lib="NIH-MGC\_118" /tissue\_type="leukocyte" /lab\_host="DH10B" /note="Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source leukocytes from anonymous pool of non-activated adult donors. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed) and 1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research genetics tracking code 027. Note: this is a NIH-MGC Library." 240 a 306 c 216 g 205 t

BASE COUNT 218 a 179 c 228 g 194 t 3 others Query Match 51.8%; Score 694.4; DB 9; Length 822; Best Local Similarity 95.9%; Pred. No. 1.5e-178; Mismatches 0; Indels 5; Gaps 5;

547 TTCAAGACCAAGCTGTGGCTGGAGCAACAAATCTGACTTTGATGTGCAACGCC 606  
Db 819 TTCAAGACCAATCTGTGGCTGGAGCAACAAATCTGAC-TTGATGTTCAAGGCC 761  
QY 607 TTCAACACAGCATTTTCCAGAGACACCTTCTCCCGAGCCAGAAAGTTCC-TGTGA 665  
Db 760 -TCAACACAGCATTTTCCAGAGACACCTTCTCCCGAGCCAGAAAGTTCCNTGTGA 702  
QY 666 TGTCAAGCTGGTGGAGAAAGCTTTGAACAGATACGAACTAACTTTCAAAACCTGTC 725  
Db 701 TGTCAAGCTGGTGGAGAAAGCTTTGAACAGATACGAACTAACTTTCAAAACCTGTC 642  
QY 726 AGTGAATGGTTCGGAATCTCTCTGAAAGTGGCGGTTTAACTGCTCATGAGCT 785  
Db 641 AGTGAATGGTTCGGAATCTCTCTGAAAGTGGCGGTTTAACTGCTCATGAGCT 582  
QY 786 GCGGCTGTGCTCAGCTGAGATCTCAAGATTGTAAGACAGCTGTGCTCCCTGCTCT 845  
Db 581 GCGGCTGTGCTCAGCTGAGATCTCAAGATTGTAAGACAGCTGTGCTCCCTGCTCT 522  
QY 846 TCCTTCTGATGGCCCTTCTCTCTCTCCCTTCCAAACAGAGGAACTCTCTACCCCAAGGA 905  
Db 521 TCCTTCTGATGGCCCTTCTCTCTCTCCCTTCCAAACAGAGGAACTCTCTACCCCAAGGA 462  
QY 906 GGTGAAGCTGCTACCACTCTGTGCCCCCGGCAATGCCCACTGGATCTTACCG 965  
Db 461 GGTGAAGCTGCTACCACTCTGTGCCCCCGGCAATGCCCACTGGATCTTACCG 402  
QY 966 AATTATGATTGAATGCTGAAGAGCTGCCAAGCTGCTGCCACCCCTCTGTTCCCT 1025  
Db 401 AATTATGATTGAATGCTGAAGAGCTGCCAAGCTGCTGCCACCCCTCTGTTCCCT 342  
QY 1026 TATTCCTGCTGTACTGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 1085  
Db 341 TATTCCTGCTGTACTGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 282  
QY 1086 GGTGTGACATTCCTCTGCT -CCCAGAGACTGCTCGCCATCCACAGATGATGG 1144  
Db 281 GGTGTGACATTCCTCTGCTCGCCATCCACAGATGATGG 222  
QY 1145 ATCTTCAGTGGTTCCTTGTGGCTCTAGGCTCTGGAGAAATGTTGAGGGGTTTA-TTTT 1203  
Db 221 ATCTTCAGTGGTTCCTTGTGGCTCTAGGCTCTGGAGAAATGTTGAGGGGTTTATTTT 162  
QY 1204 TTTTAAATAGTGTATTAAGAATATACATAGTATTTCTTCTCAAGACGTGGGGGAAA 1263  
Db 161 TTTTAAATAGTGTATTAAGAATATACATAGTATTTCTTCTCAAGACGTGGGGGAAA 102  
QY 1264 TTAATCATTTATCGAGGCTCTGATGCTGTGCTGCTGGGGTGTGATGCTGCTGCTG 1323  
Db 101 TTAATCATTTATCGAGGCTCTGATGCTGTGCTGCTGGGGTGTGATGCTGCTGCTG 42  
QY 1324 CGATGCTCTCATTAAT 1341  
Db 41 CGATGCTCTCATTAAT 24

RESULT 7  
BI911009 967 bp mRNA linear Est 16-OCT-2001  
LOCUS 603068729F1 NIH-MGC\_118 Homo sapiens cDNA clone IMAGE:5217870 5',  
DEFINITION mRNA sequence.  
ACCESSION BI911009  
VERSION BI911009.1 GI:16174524

```

FEATURES
  source
    main quality sequence stop: 717.
    Location/Qualifiers
      1. 717
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:5204103"
        /clone_lib="NIH_MGC_122"
        /lab_host="DH10B"
        /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
        Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
        anonymous pool of 24 week female lung, 16 week female
        spleen, and 20-22 week male spleens. Library is oligo-qr
        primed and directionally cloned (EcoRV site is destroyed
        upon cloning). Average insert size 1.4 kb, insert size
        range 1-3 kb. Library is normalized and enriched for
        full-length clones and was constructed by C. Gruber
        (Invitrogen). Research Genetics tracking code 026. Note:
        this is a NIH_MGC Library."
      170 a 193 c 162 g 192 t

```

```

RESULT      9
BM457321
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
1 (bases 1 to 1054)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC

```

cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM12346 row: p column: 17  
 High quality sequence stop: 660.  
 Location/Qualifiers

1. .1054  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5583616"  
 /clone\_lib="NIH\_MGC\_92"  
 /tissue\_type="embryonal carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: testis; Vector: pCMV-SPORT6; Site:1: NotI;  
 Site:2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 2.5 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."  
 257 a 308 c 237 g 251 t 1 others

FEATURES  
 source

BASE COUNT 257 a 308 c 237 g 251 t 1 others  
 ORIGIN  
 Query Match 47.2%; Score 632.8; DB 10; Length 1054;  
 Best Local Similarity 97.8%; Pred. No. 1.le-161;  
 Matches 673; Conservative 0; Mismatches 12; Indels 3; Gaps 3;

QY 379 GATATCCAGAACCTGACCTCCGCTGTCACAGCTGAGAGACTCTAAATCCAGTGACAAAG 438  
 Db 235 GATATCCAGAACCTGACCTCCGCTGTCACAGCTGAGAGACTCTAAATCCAGTGACAAAG 294  
 QY 439 TCTGCTGCTATTCACCGATTGTTGATCTCAACAAATGTTGTCACAAAGTAAAGTATCT 498  
 Db 295 TCTGCTGCTATTCACCGATTGTTGATCTCAACAAATGTTGTCACAAAGTAAAGTATCT 354  
 QY 499 GATGTGTATATCAGACAAACAGTGTAGACATGAGTCTATGAGCTTCAAGAGCAAC 558  
 Db 355 GATGTGTATATCAGACAAACAGTGTAGACATGAGTCTATGAGCTTCAAGAGCAAC 414  
 QY 559 AGTCTGTGGCTGGAGCAACAAATCTGATTTGATGTGCAACGCCCTTCAACACAGC 618  
 Db 415 AGTCTGTGGCTGGAGCAACAAATCTGATTTGATGTGCAACGCCCTTCAACACAGC 474  
 QY 619 ATTATTCAGAGACACTTCTCCCGAGCCAGAAAGTTCCTGTCATGTCAAGCTGTGTC 678  
 Db 475 ATTATTCAGAGACACTTCTCCCGAGCCAGAAAGTTCCTGTCATGTCAAGCTGTGTC 534  
 QY 679 GAGAAAGCTTTGAAACAGATACGAACCTTAAACCTTTCAAAACCTGTCAGTGATGGGTTT 738  
 Db 535 GAGAAAGCTTTGAAACAGATACGAACCTTAAACCTTTCAAAACCTGTCAGTGATGGGTTT 594  
 QY 739 CGAATCTCTCTCTGAAAGTGGCGGGTTTAACTGCTCATGAGCTGCGGCTGTGGTTC 798  
 Db 595 CGAATCTCTCTCTGAAAGTGGCGGGTTTAACTGCTCATGAGCTGCGGCTGTGGTTC 654  
 QY 799 AGCTGAGATCTCAGAGTTGTAAGACAGCTGTGCTCCCTGCTCTCTCTCTCTCTCTCTCT 858  
 Db 655 AGCTGAGATCTCAGAGTTGTAAGACAGCTGTGCTCCCTGCTCTCTCTCTCTCTCTCTCT 714  
 QY 859 CCCT 917  
 Db 715 CCCT 774  
 QY 918 TACCACCTCTGTG-CCCCCGGCGAATGCCACCACTGGATCTACCCGAAATTTATGATT 976  
 Db 775 TACCACCTCTGTGCCCCCGGCGAATGCCACCACTGGATCTACCCGAAATTTATGATT 834  
 QY \*977 AAGATTGCTGAAGAGCTGCCAACACACTGCTGCCACCCCTCTG-TTCCCTTATTGCTGCT 1035  
 Db 835 AAGATTGCTGAAGAGCTGCCAACACACTGCTGCCACCCCTCTGTTTCCCTTATTGCTGCT 894

QY 1036 TGTCACTGCTGACATTCACGGCAGAGG 1063  
 Db 895 TGTCACTGCTGACATTCACGGCAGAGG 922

RESULT 10  
 BI524586/c 913 bp mRNA linear EST 29-AUG-2001  
 LOCUS 603052009T1 NIH\_MGC\_122 Homo sapiens cDNA clone IMAGE:5201547 3',  
 DEFINITION mRNA sequence.  
 ACCESSION BI524586  
 VERSION BI524586.1 GI:15349378  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 913)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM11505 row: a column: 04  
 High quality sequence start: 22  
 High quality sequence stop: 773.  
 Location/Qualifiers

FEATURES  
 source

1. .913  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5201547"  
 /clone\_lib="NIH\_MGC\_122"  
 /lab\_host="DH10B"  
 /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;  
 Site:1: NotI; Site:2: EcoRV (destroyed); RNA source  
 anonymous pool of 24 week female lung, 16 week female  
 spleen, and 20-22 week male spleens. Library is oligo-dT  
 primed and directionally cloned (EcoRV site is destroyed  
 upon cloning). Average insert size 1.4 kb. insert size  
 range 1-3 kb. Library is normalized and enriched for  
 full-length clones and was constructed by C. Gruber  
 (Invitrogen). Research Genetics tracking code 026. Note:  
 this is a NIH\_MGC Library."  
 244 a 197 c 264 g 208 t

BASE COUNT  
 ORIGIN

Query Match 46.3%; Score 620.8; DB 10; Length 913;  
 Best Local Similarity 92.2%; Pred. No. 2e-158;  
 Matches 841; Conservative 0; Mismatches 52; Indels 19; Gaps 17;

QY 405 GTACCAGCTGAGAGACTCTAAATCCAGTGA-CAAGTCTGTCTGCTATTC-ACCGATTTT 462  
 Db 910 GTACAGCTTGAGAGACTCTTAATCCAGGACACGCTGCTGCTATTCGCCGATTC 851  
 QY 463 GATTCCTCAACAAATGTTGTACAAAGTAAGGATTCT-GATGTGTATATACAGACAAAC 521  
 Db 850 GATTCCTCAACAAATGTTGTACCAAGTAAGAAATTCACGCGGTATATACAGACATCAC 791  
 QY 522 -TGCTAG-ACATCAGGTCTATGACTTCAA---GAGCAACAGTGTGTGGCTGAGC 576  
 Db 790 TTGTGTAGAACATCAGGCTCATGACTTCAAAAGCGCAACAGTGTGTGGCCAGGAGC 731  
 QY 577 AACAAATCTGACTTTGTCATGTGCAACGCCCTTCAACACAGCATTTT-CCAGAGAGACA- 634  
 Db 730 -ACATATCGAGCTTTGTCATGTGGCAACGCCCTTCAACACAGCATTTATTTCCCAAGAGACAC 672





[illegible]

RESULT	13	LOCUS	AW514300	638 bp	nrNA	linear	EST 03-MAR-2000
DEFINITION	hd74f09.x1 NCI_CGAP_Lu28 Homo sapiens cDNA clone IMAGE:2915273 3'						
DEFINITION	similar to gb:U12959 T-CELL RECEPTOR ALPHA CHAIN C REGION (HUMAN);						
DEFINITION	nrNA sequence.						
ACCESSION	AW514300						
VERSION	AW514300.1						
KEYWORDS	EST.						
SOURCE	human.						
ORGANISM	Homo sapiens						
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
ORGANISM	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
REFERENCE	1 (bases 1 to 638)						
AUTHORS	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .						
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP).						
TITLE	Tumor Gene Index						
JOURNAL	Unpublished (1997)						
COMMENT	Contact: Robert Strausberg, Ph.D.						
COMMENT	Email: <a href="mailto:cgapb-r@mail.nih.gov">cgapb-r@mail.nih.gov</a>						
COMMENT	Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.						
COMMENT	Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life						
COMMENT	Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The						
COMMENT	I.M.A.G.E. Consortium DNA Sequencing by: Washington University						
COMMENT	Genome Sequencing Center						
COMMENT	Clone distribution: NCI-CGAP clone distribution information can be						
COMMENT	found through the I.M.A.G.E. Consortium/LLNL at:						
COMMENT	<a href="http://image.llnl.gov/image/html/iresources.shtml">image.llnl.gov/image/html/iresources.shtml</a>						
COMMENT	Seq primer: ~400p from Gibco						
COMMENT	High quality sequence stop: 380.						
FEATURES	Location/Qualifiers						
SOURCE	1..638						



```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2915273"
/clone_lib="NCI_CGAP_Lu28"
/tissue_type="fw pooled squamous cell carcinomas"
/lab_host="DH10B"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site.1: Salt;
Site.2: NotI; Cloned unidirectionally. primer: Oligo dr.
Library constructed by Life Technologies."
BASE COUNT 171 a 130 c 200 g 137 t
ORIGIN

```

```

Query Match 44.2%; Score 593.2; DB 9; Length 638;
Best Local Similarity 95.6%; Pred. No. 6.2e-151;
Matches 610; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 610 AACACAGCATATTCCAGAACACACCTTCTCCAGCCAGCCAGAAAGTCTCTGTGATGTC 669
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 638 AACACACATCATATACATGACACACACCTTACCCCTAGCCAGCAATGATCCCGCATGTC 579
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 670 AAGCTGGTCGAGAAAGCTTTGAACAGATACGAACTTCAAACTTCAAACTGTCTAGTG 729
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 578 AAGCTGGTCGAGATCGCTTTTGAACACATCAGAACTTCTTAATCTGTCTAGTG 519
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 730 ATTGGGTTCCGAATCTCTCTGAAAGTGGCGGGTTTAACTGTCTCATGAGCTGGG 789
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 518 ATTGGGTTCCGAATCTCTCTCTGACAGTGGCGGGTTTAACTGTCTCATGAGCTGGG 459
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 790 CTGTGCTCCAGCTGATCTCAAGATGTAAGACAGCTGTCTCCCTCGCTTCTTCT 849
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 458 CTGTGCTCCAGCTGATCTCAAGATGTAAGACAGCTGTCTCCCTCGCTTCTTCT 399
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 850 CTGCAATGCCCTCTTCTCCCTCTCCAAACAGAGGAATCTCTTACCCCGGAGAGTG 909
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 398 CTGCAATGCCCTCTTCTCCCTCTCCAAACAGAGGAATCTCTTACCCCGGAGAGTG 339
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 910 AAAGCTGTACCACTCTGTGCCCCCGGCAATGCCAACTGGATCTTACCCGAAT 969
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 338 AAAGCTGTACCACTCTGTGCCCCCGGCAATGCCAACTGGATCTTACCCGAAT 279
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 970 TATGATTAAGATGCTGAAGAGTGCACAACTGCTGCACCCCTCTGTTCCTTAT 1029
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 278 TATGATTAAGATGCTGAAGAGTGCACAACTGCTGCACCCCTCTGTTCCTTAT 219
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1030 GCTGCTGTCTACGCTGATCTCAGATTCAGGCGAGGCAAGGCTGCTGACGCTCCCTGGCT 1089
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 218 GCTGCTGTCTACGCTGATCTCAGATTCAGGCGAGGCAAGGCTGCTGACGCTCCCTGGCT 159
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1090 GTGCACATTCCTCTCTGCTCCCGAGAGTGCCTCCGCACTCCACAGATGATGATCTT 1149
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 158 GTGCACATTCCTCTCTGCTCCCGAGAGTGCCTCCGCACTCCACAGATGATGATCTT 99
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1150 CAGTGGTCTCTTGGGCTCTAGTCTGGAGATGTTGTGAGGGGTTTATTTTATTTA 1209
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 98 CAGTGGTCTCTTGGGCTCTAGTCTGGAGATGTTGTGAGGGGTTTATTTTATTTA 39
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1210 ATAGTGTTCATAAAGAAATACATAGTATCTTCTTCTC 1247
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38 ATAGTGTTCATAAAGAAATACATAGTATCTTCTTCTC 1.
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 14
LOCUS A1351789/c 749 bp mRNA linear EST 13-FEB-1999
DEFINITION qt09c03.x1 NCI_CGAP GC4 Homo sapiens CDNA clone IMAGE:1947076 3'
similar to gb:M12959 T-CELL RECEPTOR ALPHA CHAIN C REGION (HUMAN).;
mRNA sequence.
ACCESSION A1351789
VERSION A1351789.1 GI:4088995
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapsh-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1496 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 466.
FEATURES
source
1. 749
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1947076"
/clone_lib="NCI_CGAP_GC4"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT 198 a 176 c 205 g 167 t 3 others
ORIGIN

```

```

Query Match 43.5%; Score 584; DB 9; Length 749;
Best Local Similarity 92.0%; Pred. No. 2.1e-148;
Matches 671; Conservative 0; Mismatches 48; Indels 10; Gaps 5;

QY 614 ACAGCATTTTCCAGAGACACCTTCTCCAGCCAGCAAGTCTCTGTGATGCAAGC 673
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 749 ACAGCATTTTCCAGAGACACCTTCTCCAGCCAGCAAGTCTCTGTGATGCAAGC 693
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 674 TGGTCGAGAAAGCTTTGAACAGATACAGAACTTCAAACTTCAAACTGTGATG 733
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 692 TGGTCGAGAAAGCTTTGAACAGATACAGAACTTCAAACTTCAAACTGTGATG 639
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 734 GGTTCGAAATCTCTCTCTGAAAGTGGCGGGTTTAACTGCTCATGACGCTGGGCTGT 793
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 638 GGTTCGAAATCTCTCTCTGAAAGTGGCGGGTTTAACTGCTCATGACGCTGGGCTGT 579
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 794 GGTCCAGCTGAGATCTGCAAGATGTAAGACAGCCTGTCTCCCTCGCTTCTCTCTGC 853
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 578 GTCCAGCTGAGATCTGCAAGATGTAAGACAGCCTGTCTCCCTCGCTTCTCTCTGC 519
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 854 ATTGCCCTCTCTCTCTCCAAACAGAGGAACTCTTACCCCAAGAGGTGAAG 913
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 518 ATTGCCCTCTCTCTCTCCAAACAGAGGAACTCTTACCCCAAGAGGTGAAG 459
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 914 CTGCTACCACTCTGTGCCCCCCCCCGCAATGCCAACTGGATCTTACCCGAATATG 973
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 458 CTGCTACCACTCTGTGCCCCCCCCCGCAATGCCAACTGGATCTTACCCGAATATG 399
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 974 ATTAAGATTTGTAAGAGTGCCTAAACACTGCTGCCACCCCTCTGTTCCTTATGCTG 1033
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 398 ATTAAGATTTGTAAGAGTGCCTAAACACTGCTGCCACCCCTCTGTTCCTTATGCTG 339
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1034 CTGTGCTGCTGCTGACATTCACGGCAGAGGCAAGGC-TGCTGACGCTCCCTGGCTGTG 1092
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```





PA (BOEF ) BOEHRINGER MANNHEIM GMBH.

XX Schendel D;

XX WPI; 1998-053442/06.

DR P-PSDB; AAW47588.

XX Human T-cell receptor nucleic acids and poly:peptide(s) - for  
PT diagnosis or therapy, especially of renal cell carcinoma

XX Example 1; Pages 11-13; 30pp; German.

XX The present sequence encodes the alpha-chain of a human T-cell  
CC receptor (TCR), which can be used in the diagnosis, monitoring,  
CC prevention and therapy of a tumour disease, specifically renal  
CC cell carcinoma.

XX Sequence 1341 BP; 331 A; 365 C; 294 G; 351 T; 0 other;

Query Match 100.0%; Score 1341; DB 19; Length 1341;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGGCAAGTGGCGAGAGTGATCGTGTTCCTGACCCCTGAGTACTTTGAGCCTTGTCTAAG 60  
Db 1 atgaggaagtggcgagagtgcgtgttcctgacccctgagtaactttgagccttgctgaag 60  
QY 61 ACCACCCAGCCATCCATGACTCATATGAAGGACAAAGAGTGAACATAACCTGTAGC 120  
Db 61 accaccagcccatccatccatggactcatatgaagagacaagaagtgaacataacctgtagc 120  
QY 121 CACAACAACATTGCTACAAATGATTATACAGTGTGTACCAACAGTTTCCAGGCCCAAGGA 180  
Db 121 cacaacaacattgctacaaatgattatcacgtgtgtacacagtttccagGCCCAAGGA 180  
QY 181 CCAGCATTTATTTCAGGATACAGACAAAGATTACAAAGAGTGGCCCTCCCTGTTT 240  
Db 181 ccagcatttatttcaggatcacagacaaagattacaaagagtgccctccctgttt 240  
QY 241 ATCCCTGCGACAGAAAGTCCAGCACTCTGAGCCTGCGCCGGGTTTCCCTGAGCGACACT 300  
Db 241 atccctgcgacagaaagtccagcaactctgagcctgcccgggttccctgagcgacact 300  
QY 301 GCTGTGTACTACTGCTCGTGGTGTCTGCAAGGCAACTGACCTTTGGATCTGGGACA 360  
Db 301 gctgtgtactactgctcgtgggtgtctgcaaggcaactgacctttggatctgggaca 360  
QY 361 CAATTGACTGTTTACCTGATATCCAGAACCCCTGACCCCTGACCTGCTGACAGCTGAGAC 420  
Db 361 caattgactgttttacctgatccagaaacctgacctgacctgacctgacctgacctgacct 420  
QY 421 TCTAATCCAGTGACAAGTCTGTCTGCTATTCACCCGATTTGATTTCAACAAATGTG 480  
Db 421 tctaattccagtacaaagtctgtctgctatccaccgattttgatttctcaacaaatgtg 480  
QY 481 TCACAAAGTAAGATTTCTGATGTATATACAGACAAACTGTGCTAGACATGAGTCT 540  
Db 481 tcacaaagtaagatttctgattgtatatacacagacaaaactgtgctagacatgagctct 540  
QY 541 ATGGACTTCAAGAGCAACAGTCTGTGCGCTGGAGCACAAATCTGACTTTGCAATGTGCA 600  
Db 541 atggacttcaagagcaacagtctgtgctggagcacaaatctgactttgcaatgtgca 600  
QY 601 AAGCCCTTCAACACAGCATTTATTCAGAGACACTTCTTCCCGAGCCCAAGATTC 660  
Db 601 aagcccttcaacacagcatttattccagaagacacacttctcccgagcccaagattcc 660  
QY 661 TGTGATGTCAGCTGGTTCGAGAAAGCTTTGAAACAGATACGAACCTTAACCTTCARAAC 720  
Db 661 tgtgatgtcaagctggtcgagaaagctttgaaacagatcacgaaccttaaccttcaaac 720  
QY 721 CTGTCACTGATTGGGTTCGCAATCCCTCCTGAAAGTGGCGGGTTTAAATCTGCTCATG 780

Db 721 ctgtcagtgattgggttcgaatccctccctgaaagtggccgggtttaatctgctcag 780  
QY 781 ACGCTGCGGTGGTGGTCCAGCTGAGATCTCAAGATTGTAAGACAGCCTGTGCTCCCTCG 840  
Db 781 acgctgcggtgtgttcacagctgagatctgcaagattgtaagacagcctgtgctccctcg 840  
QY 841 CTCCTTCTGCTGCAATGCCCTCTCTTCCCTCTCTCAACAGAGAGGAACTCTCTTACCC 900  
Db 841 ctctctctctgcatgtccctcc 900  
QY 901 AAGGAGTGAAGCTGCTACACCTCTGTGCTGCCCGGCAATGCCCAACTGGATCCT 960  
Db 901 aaggagtgaaagtgtctaccacctctgtcccccccggaatgcccacacactggatcct 960  
QY 961 ACCCCAAATTTATGATTGAAGTGTGAAGAGCTGCCAAACACACTGCTGCCACCCCTCTGT 1020  
Db 961 accccaatttatgatctaagattgtgaagactgtccaaacactgtgtccacccctctgt 1020  
QY 1021 TCCCTTATTGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080  
Db 1021 tcccttattgtctgttgcactgctgacattcaccgagaggaagagctgtgctgagcct 1080  
QY 1081 CCCCTGGCTGTCACATTCCT 1140  
Db 1081 cccctggctgtgcacattccct 1140  
QY 1141 ATGGATCTTCAGTGGGTCTCTTGGGCTCTAGGCTCTGAGAAATGTTGTGAGGGTGTAT 1200  
Db 1141 atggatcttcagtggttctcttggctctaggtctctgagaaagtgtgtgaggggttat 1200  
QY 1201 TTTTCTTAAATAGTGTTCATAAAGAAATACATAGTATTCTTCTCTCTCTCTCTCTCTCT 1260  
Db 1201 ttttctttaaagtgctcataaagaaatacatagattctctctctctctctctctctctct 1260  
QY 1261 AAATATCTCATTTATCGAGGCCCTGCTATGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1320  
Db 1261 aaattatctcattatcgagggccctgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1320  
QY 1321 TGGCGATGCTTTCATTAAT 1341  
Db 1321 tgcgagtccttattataaat 1341

RESULT 2

AAS22830 standard; CDNA; 1520 BP.

AC AAS22830;

XX 24-OCT-2001 (first entry)

DE Human cDNA encoding a novel human protein #396.

XX Human; novel protein; ss; Antianemic; osteopathic; antiinflammatory;  
immunomodulatory; cytosolic; neuroprotective; vulnary; nootropic;  
KW anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;  
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;  
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
KW parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
KW tissue regeneration; immune disorder.

OS Homo sapiens.

XX WO200155437-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US02623.

XX 25-JAN-2000; 2000US-0491404.

XX (HYSE-) HYSEQ INC.

PA







us-08-881-509-1.rng

Thu Jun 6 10:09:04 2002

PR	31-JAN-2000;	2000US-0179065.	PR	02-OCT-2000;	2000US-0237037.
PR	04-FEB-2000;	2000US-0180628.	PR	02-OCT-2000;	2000US-0237038.
PR	24-FEB-2000;	2000US-0184664.	PR	02-OCT-2000;	2000US-0237039.
PR	02-MAR-2000;	2000US-0186350.	PR	13-OCT-2000;	2000US-0239935.
PR	16-MAR-2000;	2000US-0189874.	PR	13-OCT-2000;	2000US-0239937.
PR	17-MAR-2000;	2000US-0190076.	PR	20-OCT-2000;	2000US-0240960.
PR	18-APR-2000;	2000US-0198123.	PR	20-OCT-2000;	2000US-0241221.
PR	19-MAY-2000;	2000US-0205515.	PR	20-OCT-2000;	2000US-0241785.
PR	07-JUN-2000;	2000US-0209467.	PR	20-OCT-2000;	2000US-0241786.
PR	28-JUN-2000;	2000US-0214886.	PR	20-OCT-2000;	2000US-0241808.
PR	30-JUN-2000;	2000US-0215135.	PR	20-OCT-2000;	2000US-0241809.
PR	07-JUL-2000;	2000US-0216647.	PR	20-OCT-2000;	2000US-0241826.
PR	07-JUL-2000;	2000US-0216880.	PR	01-NOV-2000;	2000US-0244617.
PR	11-JUL-2000;	2000US-0217487.	PR	08-NOV-2000;	2000US-0246474.
PR	11-JUL-2000;	2000US-0217496.	PR	08-NOV-2000;	2000US-0246475.
PR	14-JUL-2000;	2000US-0218290.	PR	08-NOV-2000;	2000US-0246476.
PR	14-JUL-2000;	2000US-0220963.	PR	08-NOV-2000;	2000US-0246477.
PR	26-JUL-2000;	2000US-0220964.	PR	08-NOV-2000;	2000US-0246478.
PR	14-AUG-2000;	2000US-0224518.	PR	08-NOV-2000;	2000US-0246523.
PR	14-AUG-2000;	2000US-0224519.	PR	08-NOV-2000;	2000US-0246524.
PR	14-AUG-2000;	2000US-0225213.	PR	08-NOV-2000;	2000US-0246525.
PR	14-AUG-2000;	2000US-0225214.	PR	08-NOV-2000;	2000US-0246526.
PR	14-AUG-2000;	2000US-0225266.	PR	08-NOV-2000;	2000US-0246527.
PR	14-AUG-2000;	2000US-0225267.	PR	08-NOV-2000;	2000US-0246528.
PR	14-AUG-2000;	2000US-0225268.	PR	08-NOV-2000;	2000US-0246532.
PR	14-AUG-2000;	2000US-0225270.	PR	08-NOV-2000;	2000US-0246609.
PR	14-AUG-2000;	2000US-0225447.	PR	08-NOV-2000;	2000US-0246610.
PR	14-AUG-2000;	2000US-0225757.	PR	08-NOV-2000;	2000US-0246611.
PR	14-AUG-2000;	2000US-0225758.	PR	08-NOV-2000;	2000US-0246613.
PR	14-AUG-2000;	2000US-0225759.	PR	08-NOV-2000;	2000US-0246617.
PR	18-AUG-2000;	2000US-0226279.	PR	17-NOV-2000;	2000US-0249207.
PR	22-AUG-2000;	2000US-0226681.	PR	17-NOV-2000;	2000US-0249208.
PR	22-AUG-2000;	2000US-0226686.	PR	17-NOV-2000;	2000US-0249209.
PR	22-AUG-2000;	2000US-0227182.	PR	17-NOV-2000;	2000US-0249210.
PR	23-AUG-2000;	2000US-0227009.	PR	17-NOV-2000;	2000US-0249211.
PR	30-AUG-2000;	2000US-0228924.	PR	17-NOV-2000;	2000US-0249212.
PR	01-SEP-2000;	2000US-0229287.	PR	17-NOV-2000;	2000US-0249213.
PR	01-SEP-2000;	2000US-0229343.	PR	17-NOV-2000;	2000US-0249214.
PR	01-SEP-2000;	2000US-0229344.	PR	17-NOV-2000;	2000US-0249215.
PR	01-SEP-2000;	2000US-0229345.	PR	17-NOV-2000;	2000US-0249216.
PR	05-SEP-2000;	2000US-0229509.	PR	17-NOV-2000;	2000US-0249217.
PR	05-SEP-2000;	2000US-0229513.	PR	17-NOV-2000;	2000US-0249218.
PR	06-SEP-2000;	2000US-0230437.	PR	17-NOV-2000;	2000US-0249244.
PR	06-SEP-2000;	2000US-0230438.	PR	17-NOV-2000;	2000US-0249245.
PR	08-SEP-2000;	2000US-0231242.	PR	17-NOV-2000;	2000US-0249264.
PR	08-SEP-2000;	2000US-0231243.	PR	17-NOV-2000;	2000US-0249265.
PR	08-SEP-2000;	2000US-0231244.	PR	17-NOV-2000;	2000US-0249297.
PR	08-SEP-2000;	2000US-0231413.	PR	17-NOV-2000;	2000US-0249299.
PR	08-SEP-2000;	2000US-0231414.	PR	17-NOV-2000;	2000US-0249300.
PR	08-SEP-2000;	2000US-0232080.	PR	01-DEC-2000;	2000US-0250160.
PR	08-SEP-2000;	2000US-0232081.	PR	01-DEC-2000;	2000US-0250391.
PR	12-SEP-2000;	2000US-0231968.	PR	05-DEC-2000;	2000US-0251030.
PR	14-SEP-2000;	2000US-0232397.	PR	05-DEC-2000;	2000US-0251988.
PR	14-SEP-2000;	2000US-0232398.	PR	06-DEC-2000;	2000US-02556719.
PR	14-SEP-2000;	2000US-0232399.	PR	06-DEC-2000;	2000US-0251479.
PR	14-SEP-2000;	2000US-0232400.	PR	08-DEC-2000;	2000US-0251856.
PR	14-SEP-2000;	2000US-0232401.	PR	08-DEC-2000;	2000US-0251868.
PR	14-SEP-2000;	2000US-0233063.	PR	08-DEC-2000;	2000US-0251869.
PR	14-SEP-2000;	2000US-0233064.	PR	08-DEC-2000;	2000US-0251989.
PR	14-SEP-2000;	2000US-0233065.	PR	08-DEC-2000;	2000US-0251990.
PR	21-SEP-2000;	2000US-0234223.	PR	11-DEC-2000;	2000US-0254097.
PR	21-SEP-2000;	2000US-0234274.	PR	05-JAN-2001;	2001US-0259678.
PR	25-SEP-2000;	2000US-0234997.	XX		
PR	25-SEP-2000;	2000US-0234998.	PA	(HUMA-) HUMAN GENOME SCI INC.	
PR	26-SEP-2000;	2000US-0235484.	XX		
PR	27-SEP-2000;	2000US-0235834.	PI	Rosen CA, Barash SC, Ruben SM;	
PR	27-SEP-2000;	2000US-0235836.	XX		
PR	29-SEP-2000;	2000US-0236327.	DR	WPI; 2001-483426/52.	
PR	29-SEP-2000;	2000US-0236367.	XX		
PR	29-SEP-2000;	2000US-0236368.	PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
PR	29-SEP-2000;	2000US-0236369.	PT	useful for preventing, diagnosing and/or treating cancers and	
PR	29-SEP-2000;	2000US-0236370.	PT	metastasis -	
PR	02-OCT-2000;	2000US-0236802.			



XX PS Disclosure; SEQ ID NO 41894; 3071pp + Sequence Listing; English.

CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactivated proteins or to supplement the patients' own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/haematopoietic-related diseases, especially cancers and cancer metastases of haematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/haematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAK82169 represent sequences used in the exemplification of the present invention.

XX SQ Sequence 4658 BP; 1155 A; 1156 C; 1182 G; 1165 T; 0 other;

Query Match 39.8%; Score 534; DB 22; Length 4658;  
 Best Local Similarity 97.3%; Pred. No. 1.6e-151;  
 Matches 543; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 784 CTGCGGCTGTGGTCAGCTGAGATCTGCAAGATTGTAAGACAGCGTGTGCTCCCTCGCTC 843  
 DB 776 CTTCTGTTCTTCCTCATCCAGATCTGCAAGATTGTAAGACAGCGTGTGCTCCCTCGCTC 517

QY 844 CTTCTCTGCAATGCCCTCTCTCTCCCTCCAAACAGAGGGAACCTCTCTACCCCAAG 903  
 DB 516 CTTCTCTGCAATGCCCTCTCTCTCCCTCTCCAAACAGAGGGAACCTCTCTACCCCAAG 457

QY 904 GAGGTGAAGCTGTACCACTCTGTGCCCCCCCCCGCAATGCCCAACTGGATCTCTACC 963  
 DB 456 GAGGTGAAGCTGTACCACTCTGTGCCCCCCCCCGCAATGCCCAACTGGATCTCTACC 397

QY 964 CGAATTTATGATTAAAGATTGCTGAAGAGCTGCCAAACACTGTGCGACCCCTCTGTTC 1023  
 DB 396 CGAATTTATGATTAAAGATTGCTGAAGAGCTGCCAAACACTGTGCGACCCCTCTGTTC 337

QY 1024 CTTATGCTGCTGTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1083  
 DB 336 CTTATGCTGCTGTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 277

QY 1084 CTGCTGTGCAATTCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1143  
 DB 276 CTGCTGTGCAATTCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 217

QY 1144 GATCTTCAGTGGGTTCTCTTGGGCTCTAGTCTCTGGAAGTGTGTGAGGGGTTTATTT 1203  
 DB 216 GATCTTCAGTGGGTTCTCTTGGGCTCTAGTCTCTGGAAGTGTGTGAGGGGTTTATTT 157

QY 1204 TTTTAAATAGTCTCATAAAGAAATACATAGTATTTCTTCTCAAGACGTGGGGGAAA 1263  
 DB 156 TTTTAAATAGTCTCATAAAGAAATACATAGTATTTCTTCTCAAGACGTGGGGGAAA 97

QY 1264 TTATCTCATATTCGAGGCCCTCTATGCTGTGCTGCGGCTGTTATGCTGCTGCTG 1323  
 DB 96 TTATCTCATATTCGAGGCCCTCTATGCTGTGCTGCGGCTGTTATGCTGCTGCTGCTG 37

QY 1324 CGATGCTTCATTAAT 1341  
 DB 36 CGATGCTTCATTAAT 19

RESULT 6  
 ID AAK60079  
 XX AAK60079 standard; cDNA; 1101 BP.

AC AAN60079;  
 XX 23-JUL-1991 (first entry)  
 DT Sequence encoding a polypeptide which is at least part of the alpha chain of T cell antigen receptor.  
 DE Tumour cell; diagnosis; T cell antigen receptor; ss.  
 XX Homo sapiens.  
 OS Key  
 FH CDS Location/Qualifiers  
 FT 190..249  
 FT /\*tag= a  
 FT /product= leader sequence  
 FT 250..526  
 FT /\*tag= b  
 FT /product= variable region  
 FT 527..542  
 FT /\*tag= c  
 FT /product= diversity region  
 FT 543..594  
 FT /\*tag= d  
 FT /product= joining region  
 FT 595..945  
 FT /\*tag= e  
 FT /product= constant region  
 FT 946..1005  
 FT /\*tag= f  
 FT /product= transmembrane region  
 FT 1006..1023  
 FT /\*tag= g  
 FT /product= cytoplasmic region  
 PN EP200350-A.  
 XX 05-NOV-1986.  
 XX 25-MAR-1986; 86EP-0021945.  
 XX 15-APR-1985; 85US-0723306.  
 XX (ONTA-) ONTARIO CANCER INST.  
 XX Mak TW;  
 XX WPI; 1986-292977/45.  
 XX P-PSDB; AAF60065.  
 XX Nucleic acid encoding T-cell antigen receptor polypeptide - used in identifying T-cells etc.  
 XX Claim 4; Fig 4; 23pp; English.  
 CC AAN60079 was generated from the Jurkat human leukaemia T cell line. It is contained in T cell clone pY14. The labelled nucleic acid and monoclonal or polyclonal antibodies to the polypeptides may be used to determine whether unknown cells, e.g. tumour cells, are T cells.  
 XX Sequence 1101 BP; 263 A; 314 C; 256 G; 267 T; 1 other;

Query Match 38.3%; Score 513.8; DB 7; Length 1101;  
 Best Local Similarity 96.7%; Pred. No. 9.7e-146;  
 Matches 524; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 341 TCAGCTTTGGATCTGGGACACAAATTTGACTTTTACCTGATATCCAGACCCCTGACCCCTG 400  
 DB 560 taacttttgatcaggagacagactcagcatccggccaaatccagaaacctgacctg 619

QY 401 CCGTGTACACAGCTGAGAGACTCTAAATCCAGTACAGTCTGTCTGCTATTCACCGATT 460  
 DB 620 ccgtgtaccagctgagagactcctaaatccagtcagtcgtctgtctcttaccacctg 679



Query Match	33.38;	Score	446.4;	DB	22;	Length	834;
Best Local Similarity	77.38;	Pred.	No. 2.8e-125;				
Matches	575;	Conservative	0;	Mismatches	151;	Indels	18;
Gaps	2;						
QY	79	ATGGGACTCATATGAAGGACCAAGAGTCAACATTAACCTGTAGCCACACACAACTTCTCTACA	138				
Db	91	atgttctgcaggaaagaggctgtgactcttgactgcacatatgcacacagtgatcaac	150				

Thu Jun 6 10:09:04 2002

XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;  
 PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;  
 PI Wang A, Meagher MD;  
 XX WPI; 2001-42587/45.  
 XX New polynucleotide encoding a prostate-specific protein, for  
 PT diagnosing, monitoring and treating prostate cancer in a patient and  
 PT for use in vaccines -  
 XX  
 PS Example 24; Page 523-524; 543pp; English.  
 XX The present invention describes polynucleotide sequences (I) which encode  
 CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,  
 CC and can be used in vaccine production and gene therapy. (I), (II),  
 CC antibodies to (II), fusion proteins comprising (II), and isolated  
 CC T cells prepared using (I) or (II) are used to treat cancer in a patient.  
 CC (I) and the antibodies are also used in the detection of cancer in a  
 CC patient. The cancer that is diagnosed or treated is particularly  
 CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or  
 CC (I) can be used for monitoring the progression of cancer in a patient.  
 CC (I) and (II) can also be used to improve diagnostic and therapeutic  
 CC methods for prostate cancer. They can indicate the level of metastasis  
 CC as well as the prostate volume. AAH93357 to AAH93944 and AAH01115 to  
 CC AAH01318 represent polynucleotide and amino acid sequences used in the  
 CC exemplification of the present invention.  
 XX Sequence 834 BP; 239 A; 199 C; 194 G; 201 T; 1 other;

Query Match 33.3%; Score 446.4; DB 22; Length 834;  
 Best Local Similarity 77.3%; Pred. No. 2.8e-125;  
 Matches 575; Conservative 0; Mismatches 151; Indels 18; Gaps 2;  
 QY 79 ATGAGCTCATATGAGCAGCAAGAGTGAACATTAACCTGTAGCCACACAACTGCTTACA 138  
 DB 91 atgttcgtgcaggagaaagggtgtgactctgactgcacatgatgacaccagtatcaa 150  
 QY 139 AATGATTATATCAGTGTGTACCAAGTTCCTCCAGCAGGACCCAGGATTATTATTCAC 197  
 DB 151 agttatgtctctctgtgtacagcagccagcagtggtggtggtggtggtggtggtggt 210  
 QY 198 -----AGNATACAGACAAAGATGACAAAGAGTGGCCCTCTTTATCCCTGCC 249  
 DB 211 cagggtgtcttatgacgagcaaatgcaacagaggtgcgtactcattgaattccagaag 270  
 QY 250 GACAGAAAGTCCAGCAGCTCTGAGCCTGCCCGGGTTTCCCTGAGCGACACTGCTGTG 309  
 DB 271 gcaagaaatccgcaaccttgtcatctccgcttcaacaactggggtgactcagcaatgtat 330  
 QY 310 TACTGCTCTGTG-----GGTGTCTGCAAGGCAACTGACCTTTGGATCTGGGACA 360  
 DB 331 ttctgtgcaatgagagagggcggggagagagaaactcacctttgggacaggcaact 390  
 QY 361 CAATTGACTCTTTTACTCTATATATATATATATATATATATATATATATATATATAT 420  
 DB 391 cagctaaatggaaactcaatataccagacccctgacctgacctgacctgacctgacctg 450  
 QY 421 TCTAAATCCAGTGTGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 480  
 DB 451 tctaaatccagtgacaagtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgt 510  
 QY 481 TCACAAAGTAAAGATTCTGTATGTGTATATATATATATATATATATATATATATATAT 540  
 DB 511 tcaaaagtgaaggtatctgtatgtatataccagacaaactgtgctgacatgaggtct 570  
 QY 541 ATGACTTCAAGAGCAACAGTGTGTGGCTGAGCAACAAATCTGACTTGTGCTGCTGCTG 600  
 DB 571 atgacttcaagagcaacagt 630  
 QY 601 AACGCTTCAACACAGCATTTATCCAGAGACACCTTCTCCCGCCGACCAAGATCC 660

631 aacgcttcacaaacagcattattccagagaagacatttctcccccagccagaaagtctcc 690  
 661 TGTGATGTCAGCTGGTCGAGAAAGCTTTGAAACAGATACGAACTTAACCTTCAAAAC 720  
 691 tgtgatgtcaagctgggtcgagaaagctttgaaacagatacgaacctaactttcaaac 750  
 721 CTGTGAGTCAATGGTTCGGAATCTCTCTGAAAGTGGCGGTTTAAATCTGCTCATG 780  
 751 ctgtcagtgattgggttcgaatctctctctctctctctctctctctctctctctctct 810  
 781 ACCTGCGGCTGTGGTCCAGCTGA 804  
 811 acgtcggctgtgtgtccagctga 834  
 RESULT 10  
 AAD23603 standard; cDNA; 828 BP.  
 XX AAD23603;  
 XX 26-FEB-2002 (first entry)  
 XX Human lung tumour-specific T cell receptor alpha chain cDNA.  
 XX Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
 KW antisense-therapy; vaccine; immune response; lung cancer;  
 KW T cell receptor alpha chain; ss.  
 XX Homo sapiens.  
 XX Key 1. 828  
 XX CDS /tag= a  
 XX FT /product= "Human lung tumour-specific T cell receptor  
 FT alpha chain"  
 XX PN WO200172295-A2.  
 XX PD 04-OCT-2001.  
 XX 28-MAR-2001; 2001WO-US09991.  
 XX 29-MAR-2000; 2000US-0538037.  
 XX 05-JUN-2000; 2000US-0588937.  
 XX 18-AUG-2000; 2000US-0640878.  
 XX 22-SEP-2000; 2000US-234517P.  
 XX 01-NOV-2000; 2000US-0704512.  
 XX 14-DEC-2000; 2000US-0738973.  
 XX (CORI-) CORIXA CORP.  
 XX Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;  
 PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
 XX WPI; 2001-639201/73.  
 XX P-PSDB; AAEL3849.  
 XX New human lung-specific polynucleotides and polypeptides for the  
 XX diagnosis and treatment of disease e.g. lung cancer -  
 XX Claim 1; Page 376; 378pp; English.  
 XX The invention relates to isolated lung tumour-specific proteins and  
 CC their corresponding cDNA molecules. Lung tumour-specific proteins and  
 CC their antigen-presenting cells are useful for stimulating and/or  
 CC expanding T cells specific for a tumour protein, and for inhibiting  
 CC the development of cancer. The invention also relates to a composition  
 CC useful for stimulating an immune response, and for treating cancer. The  
 CC lung tumour specific oligonucleotide is useful in gene therapy and for  
 CC diagnosis, detection and treatment of lung cancer. The present sequence  
 CC is a cDNA encoding human lung tumour-specific T cell receptor alpha  
 CC chain.

XX	Sequence	828 BP;	235 A;	208 C;	182 G;	203 T;	0 other;
QY	338	AAC	TGAC	CTTTGG	ATCTGG	GACACAA	TTTACCTGATATCAGAACCTGACC
DB	362	agctg	ctcttga	agggaa	caagaa	ctgacag	taaacccatataccagaccctgacc
QY	398	CTG	CGGNGT	TACCA	GTGAG	ACTCTTA	ATCAGTCTGTGCTATTCACCG
DB	422	ctgcc	gtgtacc	agctg	agagact	ctaaatcc	agtgacaagctgtctgctctattaccg
QY	458	ATT	TGATTC	TCAAA	CAAAAT	GTGTCA	CAAAAGTAAAGGATTCGTGTATATCACAGCA
DB	482	attttg	attctca	acaaat	gtgtc	acaaag	taaggattctgtgtatcacagaca
QY	518	AA	ACTGTG	TACAT	GTGGT	CTATGG	ACTTCAAGAGCAACAGTGCCTGGAGCA
DB	542	aaact	gtgtag	acattg	aggtct	atggact	ccaagagcaacagctgtgctgctgagca
QY	578	ACA	AATCT	GACTT	TGCAT	GTGCA	AAACGCTTCAACACAGCATTAATCCAGACACCT
DB	602	acaa	atctg	actttg	catgt	caacag	ccttcaacacagcattattccagaagacac
QY	638	TCT	TC	CCAG	CCCG	AGAA	GTTCCTGTGATGTCAAGCTGGTTCGAGAAAAGCTTTGAAACAG
DB	662	tctt	ccagcc	agaa	agttc	ctgtg	atgctcaagctggtcgagaaaagcttgaacag
QY	698	AT	ACGA	AACT	TAACT	TTCA	AAACCTGTGAGTGTGGTTCGGAATCCCTCCTGAAAG
DB	722	at	aga	accta	aaattc	caaac	ctgtcagtgattggttccgaatccctcctctgaaag
QY	758	TG	CGCG	GTTT	AAAT	CTCT	CATGACGCTGCGGCTGTGGTCCAGCTCA
DB	782	tggc	gggtt	aatctg	ctgct	catgac	ctgcgctgtggtccagctga
RESULT	11						
ID	AAV01420	standard; cDNA to mRNA; 336 BP.					
AC	AAV01420;						
DT	19-MAY-1998	(first entry)					
DE	Human T-cell receptor alpha-chain constant region encoding cDNA.						
KW	Human; T-cell receptor; alpha-chain constant region; antigen-specific; immunosuppressant; humoral; cell mediated immune response; allergy; hypersensitivity; autoimmune reaction; transplant rejection; ds.						
OS	Homo sapiens.						
Key	Location/Qualifiers						
CDS	1..336						
FT	/*tag= a						
FT	/product= "T-cell receptor alpha-chain constant region"						
FT	/note= "no stop codon"						
XX	WO9743411-A1.						
PN	20-NOV-1997.						
XX	09-MAY-1997;	97WO-JP01565.					
XX	29-MAY-1996;	96JP-0135572.					
XX	10-MAY-1996;	96JP-0116101.					
DA	(KIRI ) KIRIN BEER KK.						

Thu Jun 6 10:09:04 2002

||||| gctgtgtactactgctctggtg 325

XX PD 02-AUG-2001.  
 XX PF 25-JAN-2001; 2001WO-US02623.  
 XX PR 25-JAN-2000; 2000US-0491404.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Liu C, Drmanac RT;  
 XX DR WPI; 2001-451939/48.  
 XX PR P-PSDB; AAU14161.  
 XX PT Isolated polypeptides useful for treating anti-inflammatory diseases,  
 XX PT nervous system disorders, and for regenerating bone and cartilage.  
 XX PS Claim 1; Page 220-221; 894pp; English.  
 XX CC The invention relates to polynucleotides encoding novel human  
 XX CC proteins or their active domains. The polypeptides, polynucleotides and  
 XX CC antibodies raised against the polypeptides are used in a method of  
 XX CC treatment of a mammal and prevention of disorders caused by the aberrant  
 XX CC protein expression or activity. The polypeptides can be used as  
 XX CC molecular weight markers, food supplements, and in antibody production.  
 XX CC The polypeptides are used to identify compounds which bind to the  
 XX CC primers, for sequencing, for chromosome or gene mapping, in the  
 XX CC production of recombinant proteins, and in generating anti-sense DNA or  
 XX CC RNA and in gene therapy. Polypeptides of the invention can be used to  
 XX CC target drugs to a tumour, in assays to determine biological activity, to  
 XX CC raise antibodies/elicite an immune response, to determine quantitative  
 XX CC protein levels, as tissue markers, and to isolate receptors or ligands.  
 XX CC Polypeptides of the invention may also be useful in treating platelet  
 XX CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
 XX CC ligament and/or nerve tissue, wound healing, treating burns, promoting  
 XX CC the proliferation, differentiation and survival of stem cells, as a  
 XX CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
 XX CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
 XX CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
 XX CC fungal infection or from autoimmunity, cancer, allergy, asthma,  
 XX CC graft-versus-host disease, eczema, haemophilia, thrombosis,  
 XX CC anti-inflammatory diseases, nervous system disorders, and infection.  
 XX CC The present sequence encodes a protein of the invention.  
 XX SQ Sequence 545 BP; 164 A; 121 C; 111 G; 149 T; 0 other;  
 Query Match 24.2%; Score 325; DB 22; Length 545;  
 Best Local Similarity 100.0%; Pred. No. 2e-86;  
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGAGGCAAGTGGCAGAGTATGCTGTTCTGACCTGAGTACTTGTGAGCCTTGTGTAAG 60  
 DB 1 atgaggaagtgccagagtgatcggttctgacccctgagctactgttgagcctgtgtaag 60  
 QY 61 ACCACCCAGCCCATCTCCATGACATCATATGAAGGACAAAGATGAACATACCTGTAGC 120  
 DB 61 accaccagcccatctccatgactcatatgaaggacaaagtgaaacataacctgagc 120  
 QY 121 CACAACAACATGCTACAAATGATATATCAGTGTGTACCAACAGATTTCCTGAGCCAAAGGA 180  
 DB 121 cacaacaacattgctacaatgatatacatcagtggtgacaaagatttcccgagcacaaga 180  
 QY 181 CCAGGATTTATTTATCAAGGATACAGACAAAGATTTACAAAGAGTGGCTCCCTGTTT 240  
 DB 181 ccagattttatttattcaagatacagaacaaagtataacagagtggtgctccctgttt 240  
 QY 241 ATCCCTGCCGACAGAAAGTCCAGCACTCTGAGCCTGCCCGGGTTTCCTGTAGCGACACT 300  
 DB 241 atccctgccgacagaaagtcagcactctgagcctgcccggtttccctgtagcgacact 300  
 QY 301 GCTGTGTACTACTGCTCGTGGGTG 325

Db 301 gctgtgtactactgctctggtg 325

RESULT 13  
 AAZ47236  
 ID AAZ47236 standard; DNA; 744 BP.  
 XX AC AAZ47236;  
 XX DT 11-APR-2000 (first entry)  
 XX DE HLA-A2/fly restricted matrix TCR-alpha/c-jun fusion coding sequence.  
 XX KW Fusion protein; T-cell receptor; c-jun; leucine zipper motif; tumour; ss;  
 XX KW influenza virus matrix protein epitope; major histocompatibility complex;  
 XX KW immune system.  
 XX OS Homo sapiens.  
 XX OS Synthetic.  
 XX PN WO9960119-A2.  
 XX PD 25-NOV-1999.  
 XX PF 19-MAY-1999; 99WO-GB01583.  
 XX PR 19-MAY-1998; 98GB-0010759.  
 XX PR 29-SEP-1998; 98GB-0021129.  
 XX PA (AVID-) AVIDEX LTD.  
 XX PI Jakobsen BK, Boulter JM;  
 XX WPI; 2000-072439/06.  
 XX CC Synthetic multivalent receptor complexes used to detect Major  
 XX CC Histocompatibility Complex-peptide complexes, and delivering  
 XX CC therapeutic agents to target cells.  
 XX CC Example 4; Fig 4; 155pp; English.  
 XX CC This sequence represents the coding region for a fusion protein  
 XX CC comprising a T-cell receptor (TCR) alpha chain linked to a c-jun leucine  
 XX CC zipper motif. The TCR gene is taken from clone JM22 and is targeted to  
 XX CC an HLA-A2 influenza virus matrix protein epitope. The invention relates  
 XX CC to a synthetic multivalent complex for binding to a major  
 XX CC histocompatibility complex (MHC)-peptide complex. The multivalent TCR  
 XX CC complexes are useful for detecting MHC-peptide complexes and delivery  
 XX CC of therapeutic agents to target cells, in vivo. The specificity of the  
 XX CC TCRs enables the localization of the liposome-contained drugs to the  
 XX CC desired target site such as a tumour or virus-infected cell. This would  
 XX CC be useful in many situations and in particular against tumours because  
 XX CC not all cells in the tumour present antigens and therefore are not all  
 XX CC detected by the immune system.  
 XX SQ Sequence 744 BP; 223 A; 182 C; 174 G; 165 T; 0 other;  
 Query Match 23.0%; Score 307.8; DB 21; Length 744;  
 Best Local Similarity 83.9%; Pred. No. 4.1e-83;  
 Matches 348; Conservative 0; Mismatches 67; Indels 0; Gaps 0;  
 QY 246 TGCCGACAGAAAGTCCAGCACTCTGAGCCTGCCCGGGTTTCCTGAGCGACACTGCTGT 305  
 DB 198 tgatgcaagaaggagacagcttctccacatcactcgcgcagcctgtgtatcacagcct 257  
 QY 306 GTACTACTGCTCGTGGGTGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACACAAATT 365  
 DB 258 ctacctctgtgcaaggagcgggaagcaagaaatctcatcttggaaaagcactaaact 317  
 QY 366 GACTGTTTACCTGATATCCAGAACCTTGACCTGCCGTGTACAGCTGAGAGACTCTTAA 425

Db 318 ctctgtttaaccacaaataccagaccctgacccctgcgtgtaccagctgagagactctaa 377  
 Qy 426 ATCCAGTGACAAGTCTCTGCTCTATTCACCGATTGATTCACAAATGCTGCACA 485  
 Db 378 atccagtgacaagctctgctgctattccacgatttgcattcacaacaaatggtcaca 437  
 Qy 486 AAGTAAGGANTCTGATGTGTATATACAGACAAACTGTGTAGACATGAGGTCTATGGA 545  
 Db 438 aagtaaggaacctgagtgtatatcacagacaaactgtctagacatgaggtctatgga 497  
 Qy 546 CTTCAAGAGCAACAGTGTCTGCTGCTGAGCAACAACTCTGATGCTGCAACAGC 605  
 Db 498 ctccaagagcaacagctgctgctgctgagcaacaaactgactttgcatgtgcaaacgc 557  
 Qy 606 CTTCAACACACAGCATATTCAGAGACACACTTCTTCCCGCCAGCAAAAGTTCC 660  
 Db 558 ctccaacacagcattattccagaagacaccttcttccccagccagaaagtccc 612

RESULT 14  
 AA247268  
 ID AA247268 standard; DNA; 744 BP.  
 AC AA247268;  
 XX  
 DT 11-APR-2000 (first entry)  
 DE HLA-A2/flu restricted matrix TCR-alpha/c-jun fusion coding sequence.  
 KW Fusion protein; T-cell receptor; c-jun; leucine zipper motif; tumour; ss;  
 KW influenza virus matrix protein epitope; major histocompatibility complex;  
 XX immune system.  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO9960119-A2.  
 PD 25-NOV-1999.  
 XX  
 PF 19-MAY-1999; 99WO-GB01583.  
 XX  
 PR 19-MAY-1998; 98GB-0010759.  
 PR 29-SEP-1998; 98GB-0021129.  
 XX  
 PA (AVID-) AVIDEX LTD.  
 XX  
 PI Jakobsen BK, Boulter JM;  
 XX WPI; 2000-072439/06.  
 DR P-PSDB; AAV56076.  
 XX  
 PT Synthetic multivalent receptor complexes used to detect Major.  
 PT Histocompatibility Complex-peptide complexes, and delivering  
 PT therapeutic agents to target cells -  
 PS Example 10; Fig 21; 155pp; English.  
 XX  
 CC This sequence represents the coding region for a fusion protein  
 CC comprising a T-cell receptor (TCR) alpha chain linked to a c-jun leucine  
 CC zipper motif. The TCR gene is taken from clone JM22 and is targeted to  
 CC an HLA-A2 influenza virus matrix protein epitope. The sequence also  
 CC contains mutations in the 5' end of the gene to enhance expression of  
 CC the protein in E. coli. The invention relates to a synthetic multivalent  
 CC complex for binding to a major histocompatibility complex (MHC)-peptide  
 CC complex. The multivalent TCR complexes are useful for detecting  
 CC MHC-peptide complexes and delivery of therapeutic agents to target cells,  
 CC in vivo. The specificity of the TCRs enables the localization of the  
 CC liposome-contained drugs to the desired target site such as a tumour or  
 CC virus-infected cell. This would be useful in many situations and in  
 CC particular against tumours because not all cells in the tumour present  
 CC antigens and therefore are not all detected by the immune system.

SQ Sequence 744 BP; 223 A; 182 C; 174 G; 165 T; 0 other;  
 Query Match 23.0%; Score 307.8; DB 21; Length 744;  
 Best Local Similarity 83.9%; Pred. No. 4.1e-83;  
 Matches 348; Conservative 0; Mismatches 67; Indels 0; Gaps 0;  
 QY 246 TCCCGCAGACAAAGTCCAGCACTCTGAGCCCTGCCCCGGGTTTCCCTGAGCGACACTGCTGT 305  
 Db 198 tgatgcaagaagaagacagtctctccacatcactgcggccagcctggtgaTaccaggcct 257  
 QY 306 GTACTACTGCTCGTGGGTGGTTCGCAAGGCAACTGACCTTTGGATCTGGGACACACATT 365  
 Db 258 ctactctgtgcaggagcggaagcgaagaaatctcatcttttgaaagcactaaact 317  
 QY 366 GACTGTTTTACCTGATATCCAGAACCTTACCCTGCGCTGTACACAGCTGAGAGACTCTAA 425  
 Db 318 ctctgttaaacacaaataccagaacccctgacctgcgtgtaccagctgagagactctaa 377  
 QY 426 ATCCAGTGACAAGTCTCTGCTGCTATTCACCGATTGATTCACAAATGCTGCACA 485  
 Db 378 atccagtgacaagctctgctgctattccacgatttgcattcacaacaaatggtcaca 437  
 QY 486 AAGTAAGGANTCTGATGTGTATATACAGACAAACTGTGTAGACATGAGGTCTATGGA 545  
 Db 438 aagtaaggaacctgagtgtatatcacagacaaactgtctagacatgaggtctatgga 497  
 QY 546 CTTCAAGAGCAACAGTGTCTGCTGCTGAGCAACAACTCTGATGCTGCAACAGC 605  
 Db 498 ctccaagagcaacagctgctgctgagcaacaaactgactttgcatgtgcaaacgc 557  
 QY 606 CTTCAACACACAGCATATTCAGAGACACACTTCTTCCCGCCAGCAAAAGTTCC 660  
 Db 558 ctccaacacagcattattccagaagacaccttcttccccagccagaaagtccc 612

RESULT 15  
 AA256659  
 ID AA256659 standard; DNA; 744 BP.  
 XX  
 AC AA256659;  
 XX  
 DT 22-MAR-2000 (first entry)  
 DE TCR alpha chain and c-jun fusion protein encoding DNA - Fig.10.  
 KW T cell receptor; TCR; soluble; cytotoxic; immunostimulatory; MHC;  
 KW major histocompatibility complex; vaccine; infection; detection;  
 KW diagnosis; autoimmune disease; tumour; viral disease; cancer;  
 KW metastasis; graft rejection; ss.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO9960120-A2.  
 PD 25-NOV-1999.  
 XX  
 PF 19-MAY-1999; 99WO-GB01588.  
 XX  
 PR 19-MAY-1998; 98GB-0010759.  
 PR 29-SEP-1998; 98GB-0021129.  
 XX  
 PA (AVID-) AVIDEX LTD.  
 XX  
 PI Jakobsen BK, Bell JL, Gao GF, Willcox BE, Boulter JM;  
 XX WPI; 2000-086595/07.  
 DR P-PSDB; AAV57852.  
 XX  
 PT Refolded recombinant T-cell receptors used to detect the efficacy of T  
 PT cell vaccines -  
 XX

Example 7; Fig 10; 125pp; English.

PS The present invention describes a refolded recombinant T-cell receptor  
XX (TCR) comprising: (a) a recombinant TCR alpha or gamma chain dimerisation  
CC extracellular domain having a first heterologous C-terminal dimerisation  
CC peptide; and (b) a recombinant TCR beta or delta chain extracellular  
CC domain having a second C-terminal dimerisation peptide which is  
CC specifically heterodimerised with the first dimerisation peptide to form  
CC a heterodimerisation domain. The recombinant TCR can be used for  
CC investigating specific TCR-MHC (major histocompatibility complex)  
CC interactions, and also as a diagnostic tool to detect infection or  
CC autoimmune disease markers, or to detect the efficacy of T cell vaccines.  
CC The TCR could be used to deliver therapeutic agents to cells presenting a  
CC particular antigen, e.g. delivering a toxin to a tumour to eradicate it.  
CC The TCRs would be beneficial in treating viral diseases for which drugs  
CC exist, by allowing for the drug to be released in the near vicinity of  
CC infected cells. For cancer, the localisation in the vicinity of tumours  
CC or metastasis using the TCRs would enhance the effect of toxin or  
CC immunostimulants. In autoimmune diseases, using the TCRs,  
CC immunosuppressive drugs could be released slowly, having more local  
CC effect over a longer time-span while minimally affecting the overall  
CC immunocapacity of the subject. In the prevention of graft rejection,  
CC the effect of immunosuppressive drugs could be optimised in the same  
CC way using the TCRs. The present sequence encodes a TCR fusion protein  
CC from an example from the present invention.

XX Sequence 744 BP; 223 A; 182 C; 174 G; 165 T; 0 other;  
SQ

Query Match		23.0%;	Score 307.8;	DB 21;	Length 744;
Best Local Similarity		83.9%;	Pred. No. 4.1e-83;		
Matches 348;		Conservative 0;	Mismatches 67;	Indels 0;	Gaps 0;
QY	246	TGCCGACAGAAAGTCCAGACACTGTGACCTGCGCCGGGTTTCCTGAGCGACACTGCTGT	305		
DB	198	tgatgcaagaagagacagattctctccacatcactcgccgcccgcctgtgtatcacaggcct	257		
QY	306	GTACTACTGCTCTGCTGGTGCTTCTGCAAGGCAACTGACCTTTGGATCTGGGACACAAATT	365		
DB	258	ctacctctgtgcaggagcggaagcaaggaatctcatctcttggaaaggccactaaact	317		
QY	366	GACTGTTTATCTGATATCCAGAACCCCTGACCCCTGCGGTGACCGTGGAGACTCTAA	425		
DB	318	ctctgttaaaccaaatccagaaacctgacccctgcccgtgtaccagctgagagactctaa	377		
QY	426	ATCCAGTGACAGTCTGCTGCTATTCACCGATTTTGATTTCAAAACAAATGTGTACA	485		
DB	378	atccagtgacaagtctgtcgtctattccacgattttgattctcaacaaatgtgtcaca	437		
QY	486	AAGTAAGGATTCTGATGTATATACAGACAAACTGTGCTAGACATGAGGCTATGGA	545		
DB	438	aagtaagattctgtatgtatatacacagacaaactgtctagacatgaggtctatgga	497		
QY	546	CTTCAAGACACAGTGTGTGGCTGGAGCAACAAATCTGACTTTGCATGTGCAAAACGC	605		
DB	498	cttcaagagcaacagtgtgtgtgctgagcaacaaactgtactttgcatgtgcaaacgc	557		
QY	606	CTTCAACACAGCATTTATCCAGAGACACCTTTCTTCCCGCCAGCAAGTTCC	660		
DB	558	cttcaacaacagcattatttccagagacacaccttcttccccagcccgagaaagtcc	612		

Search completed: June 6, 2002, 07:00:20  
Job time: 3043 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 06:09:38 ; Search time 78.65 seconds  
(without alignments)  
4188.106 Million cell updates/sec

Title: US-08-881-509-1  
Perfect score: 1341  
Sequence: 1 ATGAGCAAGTGGCAGAGT.....GCCGATGCTTCATTAAAT 1341

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PCRTUS\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	448	33.4	822	1	US-07-936-267A-8
2	299.6	22.3	1372	6	5189147-2
3	264.4	19.7	990	4	US-09-082-593-9
4	247.4	18.4	804	1	US-08-416-336-1
5	57	4.3	747	4	US-09-140-084-25
6	55.8	4.2	7218	1	US-08-232-463-14
7	41.8	3.1	93	2	US-08-411-098-6
8	39.8	3.0	645	2	US-08-480-753-7
9	37.2	2.8	4691	4	US-08-591-632-43
10	37.2	2.8	6166	4	US-08-591-632-51
11	36.8	2.7	330	4	US-09-025-769B-46
12	36.4	2.7	7635	1	US-08-619-554-1
13	35.6	2.7	402	1	US-08-405-034-3
14	35.6	2.7	646	3	US-08-300-386A-2
15	35.6	2.7	646	3	US-08-931-645-2
16	35.6	2.7	646	5	PCT-US94-01258-2
17	35.6	2.7	646	5	PCT-US95-11235-2
18	35.6	2.7	714	1	US-07-971-160-1
19	35.6	2.7	714	1	US-08-336-241-1
20	35.6	2.7	714	2	US-08-465-273-1
21	35.6	2.7	714	2	US-09-119-024-1
22	35.6	2.7	714	2	US-08-417-226-1
23	35.6	2.7	714	4	US-09-196-131-1
24	35.6	2.7	754	1	US-07-971-160-5
25	35.6	2.7	754	1	US-08-336-241-5
26	35.6	2.7	754	2	US-08-465-273-5
27	35.6	2.7	754	2	US-09-119-024-5

Sequence 5, Appl  
Sequence 5, Appl  
Sequence 9, Appl  
Sequence 9, Appl  
Sequence 9, Appl  
Sequence 9, Appl  
Sequence 9, Appl  
Sequence 9, Appl  
Sequence 3, Appl  
Sequence 241, App  
Sequence 241, App  
Sequence 1, Appl  
Sequence 9, Appl  
Sequence 9, Appl  
Sequence 17, Appl  
Sequence 17, Appl  
Sequence 11, Appl

## ALIGNMENTS

## RESULT 1

US-07-936-267A-8  
; Sequence 8, Application US/07936267A  
; Patent No. 5445940  
; GENERAL INFORMATION:  
; APPLICANT: Michael B. Brenner and Harout Dersimonian  
; TITLE OF INVENTION: V12.1 T-Cell Receptor-  
; TITLE OF INVENTION: Specific Reagents For Diagnosing And Treating  
; TITLE OF INVENTION: Rheumatoid Arthritis  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dana-Farber Cancer Institute, Inc.  
; STREET: 44 Binney Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02115

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 720Kb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Personal System 2; Model 30  
; OPERATING SYSTEM: MS/DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION NUMBER: US/07/936,267A  
; FILING DATE: 26-AUG-1992  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/750,986  
; FILING DATE: 28-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HART, JULIA D.  
; REGISTRATION NUMBER: 33132  
; REFERENCE/DOCKET NUMBER: DFCI-121.1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 632-4016  
; TELEFAX: (617) 632-4012  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 822 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: no  
; ANTI-SENSE: no  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapien  
; DEVELOPMENTAL STAGE: adult  
; CELL TYPE: Human T-cell leukemia line HBM-MLT

RESULT 2  
5189147-2  
; Patent No. 5189147  
; APPLICANT: SAITO, HARUO; KRANZ, DAVID M.; ELSEN, HERMAN N.;  
; TONEGAWA, SUSUMU  
; TITLE OF INVENTION: METEROIDIMERIC T LYMPHOCYTE RECEPTOR  
; ANTIBODY  
; NUMBER OF SEQUENCES: 21  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/271,216  
; FILING DATE: 14-NOV-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 666,988  
; FILING DATE: 31-OCT-1984  
; APPLICATION NUMBER: 620,122  
; FILING DATE: 13-JUN-1984  
; SEQ ID NO: 2:  
; LENGTH: 1372  
5189147-2

Query Match 22.3%; Score 299.6; DB 6; Length 1372;  
Best Local Similarity 69.2%; Pred. No. 5.1e-79;  
Matches 429; Conservative 0; Mismatches 179; Indels 12; Gaps 1;

QY 259 TCAGACACCTCTGAGCCTGCCCGGGTTTCCTCTGAGCGACACTGCTGTACTACTGCTC 318  
Db 329 tccttcacactcggaagacccctcgctgagcgtggagcgactcggtgtgtactctgtgct 388  
QY 319 GTGGGTGGTCTTCAAGGCAACTGACCTTTGGATCTGGGACACAAATGACATGTTTACCT 378  
Db 389 gtgagcggttgcgaagtgcgtgacattggatcgccacaaagtcatgttctacca 448  
QY 379 GATATCCAGAACCTGACCTGCCGTGACGACTGAGAGACTCTAAATCCAGTGACAAG 438  
Db 449 tacatcagaacccagaaacctgtgtaccagtttaaaagatccctcgctgacagagca 508  
QY 439 TCTGTCTGCTATTCACCGATTTTGTATCTCAACAAATGTGTCAACAAGTAAGATTC 498  
Db 509 acctctgctgttaccgaacttgactcccaaatcaatgctgcgaagacatggaatct 568  
QY 499 GATGTGTATATACAGACAAACTGTGTAGACATGAGGTCTATGACTTCAAGAGCAAC 558  
Db 569 ggaacgttcatcactgacaaaactgtgtggacatgaaagctatgattcccaagagcaat 628  
QY 559 AGTGTCTGCTGCGCTGGAGCAACAAATCTGACTTTGATGTGCAAAAGCTTCAACACAGC 618  
Db 629 ggggccaattgctcctggagcaaccagacaagcttccactgccaagatatctcaagagacc 688  
QY 619 ATTATTTCCAGAGACACACTCTTCTTCCCGAGCCCGAGAAAGTTCTGTGTGATGTCAGTGGTC 678  
Db 689 a-----acgccaactaccacagttcagacgttccctgtgtatgcccacgttgacc 736  
QY 679 GAGAAAAGCTTTGAAACAGATAGCAACCTTAACTTTCAAAACCTGTCAAGTGGTTC 738  
Db 737 gagaaaagctttgaaacagatatgaacctaaactttcaaaacctgtcagttatggactc 796  
QY 739 CGAATCTCTCTCTGAAAGTGGCGGGTTPAATCTGCTCATGACGCTGGCGGTGTGGTCC 798  
Db 797 cgaatcctcctgtgaaagttagcgggatttaacctgctcatgacgtgaggtgtggtcc 856  
QY 799 AGCTGAGATGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCTCTCTGCAATTGC 858  
Db 857 agttagaggtgtgcaagactgacagagcctgactcccaagttccgtcctcctaccctcc 916  
QY 859 CCCTCTCTCTCTCTCCAAA 878  
Db 917 gctccctcttcaagccaaaa 936

FEATURE:  
; NAME/KEY: T cell receptor '-chain-encoding cDNA sequence  
; LOCATION: 1-822 Encoded amino acids -20 to -1 are  
; OTHER INFORMATION: predicted to be the leader sequence; encoded amino acids 1-95  
; OTHER INFORMATION: predicted to be the variable region; encoded amino acids 96-1  
; OTHER INFORMATION: predicted to be the joining region; encoded amino acids 112-2  
; OTHER INFORMATION: predicted to be the constant region; encoded amino acids 229  
; OTHER INFORMATION: are predicted to be the transmembrane region; and encoded am  
; OTHER INFORMATION: acids 249-253 are predicted to be the cytoplasmic region.  
; PUBLICATION INFORMATION:  
; AUTHORS: Sim et al.  
; TITLE: Primary structure of human T-cell receptor '-chain  
; JOURNAL: Nature  
; VOLUME: 312  
; PAGES: 771-775  
; DATE: 1984  
; US-07-936-267A-8

Query Match 33.4%; Score 448; DB 1; Length 822;  
Best Local Similarity 77.8%; Pred. No. 4.3e-123;  
Matches 557; Conservative 0; Mismatches 150; Indels 9; Gaps 1;

QY 98 AGAAGTGACATACCTGTAGCCCAACACACATGTGACAAATGATTATATACAGTGGT 157  
Db 107 AGGATGTGACCTTGGACTGTGTGTATGAACCCGTGACTACTATTACTTATCTGGT 166  
QY 158 ACCAAGAGTTTCCAGCCCAAGGACACGATTTATTATTATCAAGGATAC-----AAGA 208  
Db 167 ACAAGCAACCCAAAGTGGAATGGTTTCTTATTCTGCGGAACCTTTTGTATGAGC 226  
QY 209 CAAAAGTTACAAAGAGTGGCTCCCTGTTTATCCCTGCCGACAGAAAGTCCAGCACT 268  
Db 227 AAAATGAATTAAGTGTGCGTATFTCTTGAACCTCCAGAAATCCACAGTTCCTTCACT 286  
QY 269 TGAGCCTGCCCGGGTTTCCTGAGCGACACACTGCTGTGTACTGCTGCTGCTGCTGCTG 328  
Db 287 TCACCATCAGACCTCACAAGTGTGGACTGACGACTATATCTGCTGCTGCTGCTGCA 346  
QY 329 CTGCAAGGCAACTGACCTTTGGATCTGGGACACAAATGACTCTTTTACTGATATCCAGA 388  
Db 347 GTGCTTCCAAGATAATCTTTGGATCAGGACACAGACTCAGCATCCGCGCAATATCCAGA 406  
QY 389 ACCCTGACCTGCCGTGTACAGTGTGAGAGACTCTAAATCCAGTCAAGTCTGTCTGCC 448  
Db 407 ACCCTGACCTGCCGTGTACAGTGTGAGAGACTCTAAATCCAGTCAAGTCTGTCTGCC 466  
QY 449 TATTCACCGATTTTGATTTCAACAAATGTGTCAAAAGTAAAGGATTCTGTATGTCTATA 508  
Db 467 TATTCACCGATTTTGATTTCAACAAATGTGTCAAAAGTAAAGGATTCTGTATGTCTATA 526  
QY 509 TCACAGCAAAACTGTCTAGACATGAGTCTATGGACTTCAAGACCAACAGTGTGTGG 568  
Db 527 TCACAGCAAAACTGTCTAGACATGAGTCTATGGACTTCAAGACCAACAGTGTGTGG 586  
QY 569 CTTGAGCAACAAATCTGACTTTGCATGTGCAAAAGCTTCAACACAGCATTAATTCAG 628  
Db 587 CTTGAGCAACAAATCTGACTTTGCATGTGCAAAAGCTTCAACACAGCATTAATTCAG 646  
QY 629 AAGACACCTTCTTCCCGAGCCCAAGAGTTCCTGTGATGTCAAGCTGGTGGTGAGAAAGCT 688  
Db 647 AAGACACCTTCTTCCCGAGCCCAAGAGTTCCTGTGATGTCAAGCTGGTGGTGAGAAAGCT 706  
QY 689 TTGAACAGATACGAACCTTAACTTTCAAAACCTGTGCTGAGTGGTTCCTGATCCTCC 748  
Db 707 TTGAACAGATACGAACCTTAACTTTCAAAACCTGTGCTGAGTGGTTCCTGATCCTCC 766  
QY 749 TCTGAAAGTGGCGGGTTTAACTGTCTATGACGCTGCGGCTGTGCTCCAGCTGA 804  
Db 767 TCTGAAAGTGGCGGGTTTAACTGTCTATGACGCTGCGGCTGTGCTCCAGCTGA 822

RESULT 3  
US-09-082-593-9  
; Sequence 9, Application US/09082593

```

; Sequence 1, Application US/08416336
; Patent No. 5807714
; GENERAL INFORMATION:
; APPLICANT: Ishizaka, Kimishige
; APPLICANT: Ishii, Yasuyuki
; TITLE OF INVENTION: METHOD OF PRODUCTION OF ANTIGEN-SPECIFIC
; TITLE OF INVENTION: GLYCOSYLATION INHIBITING FACTOR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/416.336

```

[illegible]

A



```
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411,098
; FILING DATE: 27-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CAROL M. GRUPPI
; REGISTRATION NUMBER: 37,341
; REFERENCE/DOCKET NUMBER: 2026-4188
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93
; TYPE: NUCLEOTIDE
; STRANDEDNESS: DOUBLE
; TOPOLOGY: UNKNOWN
; US-08-411-098-6

Query Match 3.1%; Score 41.8; DB 2; Length 93;
Best Local Similarity 65.6%; Pred. No. 0.0029;
Matches 61; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 304 GTGTACTACTGCTCTGGTGGTTCGCAAGGCAACTGACCTTTGGATCTGGGACACAA 363
Db 1 GAGTACTTCTGCTGCTGGTGGTCCACCGGTACCACTTCTATTTGGGACAGGCAAGT 60

QY 364 TTGACGTGTTTACCTCATATCCAGAACCTCGAC 396
Db 61 TTGACGGTCATTCGAATATCCAGAACCTCGAC 93

RESULT 8
US-08-480-753-7
; Sequence 7, Application US/08480753
; Patent No. 5830675
; GENERAL INFORMATION:
; APPLICANT: Targan M.D., Stephan R.
; TITLE OF INVENTION: METHODS FOR SELECTIVELY DETECTING
; TITLE OF INVENTION: PERINUCLEAR ANTI-NEUTROPHIL CYTOPLASMIC ANTIBODY OF
; TITLE OF INVENTION: ULCERATIVE COLITIS, PRIMARY SCLEROSING CHOLANGITIS, OR
; TITLE OF INVENTION: TYPE I AUTOIMMUNE HEPATITIS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wendy A. Whiteford, Esq.
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,753
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Whiteford, Wendy A.
; REGISTRATION NUMBER: 36,964

; REFERENCE/DOCKET NUMBER: P07 33571
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 622-7700
; TELEFAX: (213) 489-4210
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 645 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Gut-associated lymphoid
; CELL TYPE: Lymphocyte
; IMMEDIATE SOURCE:
; CLONE: 5-4
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..645
; OTHER INFORMATION: /codon_start= 1
; OTHER INFORMATION: /product= "Kappa Light Chain of ANCA associated
; OTHER INFORMATION: with Ulcerative Colitis"
; FEATURE:
; NAME/KEY: misc_rna
; LOCATION: 1..3
; OTHER INFORMATION: /label= N-TerminalTag
; FEATURE:
; NAME/KEY: misc_rna
; LOCATION: 4..285
; OTHER INFORMATION: /label= VKSEGMENT
; OTHER INFORMATION: /note= "VKSEGMENT" refers to Variable Segment of
; OTHER INFORMATION: the Kappa Light Chain"
; FEATURE:
; NAME/KEY: misc_rna
; LOCATION: 286..327
; OTHER INFORMATION: /label= JK
; OTHER INFORMATION: /note= "JK" refers to Joining Segment of the
; OTHER INFORMATION: Kappa Light Chain"
; FEATURE:
; NAME/KEY: misc_rna
; LOCATION: 328..645
; OTHER INFORMATION: /label= CK
; OTHER INFORMATION: /note= "CK" refers to Constant Segment of the
; OTHER INFORMATION: Kappa Light Chain"
; FEATURE:
; NAME/KEY: misc_rna
; LOCATION: 4..66
; OTHER INFORMATION: /label= FR1
; OTHER INFORMATION: /note= "FR1" refers to Framework Region 1"
; FEATURE:
; NAME/KEY: misc_rna
; LOCATION: 67..102
; OTHER INFORMATION: /label= CDR1
; OTHER INFORMATION: /note= "CDR1" refers to Complementarity
; OTHER INFORMATION: Determining Region 1"
; FEATURE:
; NAME/KEY: misc_rna
; LOCATION: 103..147
; OTHER INFORMATION: /label= FR2
; OTHER INFORMATION: /note= "FR2" refers to Framework Region 2"
; FEATURE:
; NAME/KEY: misc_rna
; LOCATION: 148..168
; OTHER INFORMATION: /label= CDR2
; OTHER INFORMATION: /note= "CDR2" refers to Complementarity
; OTHER INFORMATION: Determining Region 2"
; FEATURE:
; NAME/KEY: misc_rna
; LOCATION: 169..264
; OTHER INFORMATION: /label= FR3
```

OTHER INFORMATION: /note= "FR3" refers to Framework Region 3"  
FEATURE:  
NAME/KEY: misc\_rna  
LOCATION: 265..294  
OTHER INFORMATION: /label= CDR3  
OTHER INFORMATION: /note= "CDR3" refers to Complementarity  
OTHER INFORMATION: Determining Region 3"  
FEATURE:  
NAME/KEY: misc\_rna  
LOCATION: 295..327  
OTHER INFORMATION: /label= FR4  
OTHER INFORMATION: /note= "FR4" refers to Framework Region 4"  
US-08-480-753-7  
Query Match 3.0%; Score 39.8; DB 2; Length 645;  
Best Local Similarity 45.6%; Pred. No. 0.031;  
Matches 140; Conservative 0; Mismatches 167; Indels 0; Gaps 0;  
QY 70 CCCATCTCCATGGACTCATATGAAGGACAAAGAGTGAACATAACCTGTAGCCACAAAC 129  
Db 19 CCAGGACCCCTGCTTTGCTCCAGGGAAGAGCCACCTCTCTCTGCGAGGGCCAGTCAG 78  
QY 130 ATTGCTACAAATGATTATATACCTGGTGTACCAACAGTTTCCAGCCCAAGGACCAAGATT 189  
Db 79 GGTGTAGCAGCGCTCCTTAGCTGCTGATATCAGCAGAAAGCTGGCCAGGCTCCAGGCTC 138  
QY 190 ATTATTCAGGATACAGACAAAGTTTACAAAGAGTGGCCCTCCCTGTTTATCCCTGCC 249  
Db 139 CTCATCTATGTGTCATCAGAGGGCCACTGGCATCCAGACAGAGTTCACTGGCAGTGGG 198  
QY 250 GACAGAAAGTCCAGCACTCTGAGCCTGCCCGGGTTTCCCTGAGCGACACTGTGTGTAC 309  
Db 199 TCTGGACAGACTTCACTCTCACATCACAGACTGGAGCTGAAGATTITGCACTGTAT 258  
QY 310 TACTGCCCTGCTGGGTGTTCTGCAAGGCACTGACCTTTGGATCTGGGACACAACTGACT 369  
Db 259 TACTGTACAGTAGTGTAGTCCAGGAGTTCATCTTCCGCGCTGGGACCAAGTGGAT 318  
QY 370 GTTTTAC 376  
Db 319 CTCAAAC 325  
RESULT 9  
US-08-591-632-43  
Sequence 43, Application US/08591632  
Patent No. 6261558  
GENERAL INFORMATION:  
APPLICANT: Barbas, Carlos F.  
APPLICANT: Burton, Dennis R.  
APPLICANT: Lerner, Righard A.  
TITLE OF INVENTION: SYNTHETIC HUMAN NEUTRALIZING MONOCLONAL  
TITLE OF INVENTION: ANTIBODIES TO HUMAN IMMUNODEFICIENCY VIRUS  
NUMBER OF SEQUENCES: 92  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: The Scripps Research Institute, Office of  
ADDRESSEE: Patent Counsel  
STREET: 10550 No. 6261558th Torrey Pines Road, TPC 8  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/591.632  
FILING DATE: 19-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/11907

FILING DATE: 19-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/308,841  
FILING DATE: 19-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/233,619  
FILING DATE: 26-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/139,409  
FILING DATE: 19-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitting, Thomas  
REGISTRATION NUMBER: 34,163  
REFERENCE/DOCKET NUMBER: TSRI 332.3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 784-2937  
TELEFAX: (619) 784-9399  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4691 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
US-08-591-632-43  
Query Match 2.8%; Score 37.2; DB 4; Length 4691;  
Best Local Similarity 46.5%; Pred. No. 0.52;  
Matches 120; Conservative 0; Mismatches 138; Indels 0; Gaps 0;  
QY 70 CCATCTCCATGGACTCATATGAAGGACAAAGAGTGAACATAACCTGTAGCCACAAAC 129  
Db 2692 CCAGGACCCCTGCTTTGCTCCAGGGAAGAGCCACCTCTCTCTGCGAGGGCCAGTCAC 2751  
QY 130 ATTGCTACAAATGATTATATACCTGGTGTACCAACAGTTTCCCGAGCAAGGACCAAGATT 189  
Db 2752 AGTGTAGCAGGCTTACTTACCTGGTGTACCAACAGTTTCCCGAGCAAGGACCAAGATT 2811  
QY 190 ATTATTCAGGATACAGACAAAGTTTACAAAGAGTGGCCCTCCCTGTTTATCCCTGCC 249  
Db 2812 CTCATCTATGTTATCTCCAGAGGGCCACTGGCATCCAGACAGAGTTCACTGGCAGTGGG 2871  
QY 250 GACAGAAAGTCCAGCACTCTGAGCCTGCCCGGGTTTCCCTGAGCGACACTGTGTGTAC 309  
Db 2872 TCTGGACAGACTTCACTCTCACATCACAGACTGGAGCTGAAGATTITGCACTGTAT 2931  
QY 310 TACTGCCCTGCTGGGTGGT 327  
Db 2932 TACTGTACAGAGTATGTT 2949  
RESULT 10  
US-08-591-632-51  
Sequence 51, Application US/08591632  
Patent No. 6261558  
GENERAL INFORMATION:  
APPLICANT: Barbas, Carlos F.  
APPLICANT: Burton, Dennis R.  
APPLICANT: Lerner, Righard A.  
TITLE OF INVENTION: SYNTHETIC HUMAN NEUTRALIZING MONOCLONAL  
TITLE OF INVENTION: ANTIBODIES TO HUMAN IMMUNODEFICIENCY VIRUS  
NUMBER OF SEQUENCES: 92  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: The Scripps Research Institute, Office of  
ADDRESSEE: Patent Counsel  
STREET: 10550 No. 6261558th Torrey Pines Road, TPC 8  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: James F. Haley, Jr., Esq. c/o Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10021

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/025,769B
FILING DATE: 18-FEB-1998
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 95 11 3021.0
FILING DATE: 18-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: James F. Haley, Jr., Esq.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: MORPHO/5
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)596-9000
TELEFAX: (212)596-9090
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 330 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "synthetic gene"
FEATURE:
NAME/KEY: CDS
LOCATION: 1..330
OTHER INFORMATION: /product= "v kappa 3"
US-09-025-769B-46

```

Query Match	2.7%	Score 36.8	DB 4	Length 330
Best Local Similarity	47.1%	Pred. No. 0.17		
Matches 113	Conservative 0	Mismatches 127	Indels 0	Gaps 0
QY	77	CCATGGACTCATATGAAGGACAAGTGAACATATACCTGTAGCCACACACATTGCTA	136	
Db				
QY	29	CCCTGAGCCTGTC CCGGCGAAGCTGCGACCTGAGCTGCAGAGGAGCCAGAGCGTGA	88	
Db				
QY	137	CAAAATGATTTATCATCGTGTGTACCAACAGTTTCCCAGCCCAAGSACCCAGATTATTATTATTC	196	
Db				
QY	89	GCAGCAGCATATCTGGCGTGTGTACCAAGCAGAAACAGGCTCAAGCAGCGGCTCTATTAAATTT	148	
Db				
QY	197	AAGGATACAAGACAAAAGTTTACAACCAAGATGGCCCTCCCTGTTTATCCCTGCCGACAGAA	256	
Db				
QY	149	ATGGCGCAGCAGCGCGTGCACACTGGGGTCCCGGCGGTTTTAGCGGCTCTGGATCCGGCA	208	
Db				
QY	257	AGTCACGACACTCTGAGCGCTGCCCGGGTTTCCCTGAGCAGCACTGCTGTGTACTACTGCC	316	
Db				
QY	209	CGGAATTTTACCCTCCACCATTTAGCAGCGCTGGAACCTTGAAGACTTTTCGGGTGTTATTATTC	268	
Db				

RESULT 12  
US-08-619-554-1  
; ; Sentence 1, Application US/08619554  
; Patent No. 5821353  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS, Cameron M.  
; APPLICANT: CHREBET, Gary L.  
; APPLICANT: CLEMAS, Joseph  
; APPLICANT: EL-SHERBINI, Mohammed  
; APPLICANT: FOOR, Forrest  
; APPLICANT: KAHN, Jennifer  
; APPLICANT: KELLY, Rosemarie, - PA

```

RESULT 14
US-08-300-386A-2
; Sequence 2, Application US/08300386A
; Patent No. 5667988
; GENERAL INFORMATION:
; APPLICANT: Barbas, Carlos F, III
; APPLICANT: Burton, Dennis R
; APPLICANT: Lerner, Richard A
; TITLE OF INVENTION: METHODS FOR IMPROVING THE EFFICIENCY OF USING UNIVERSAL
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute
; STREET: 10666 No. 5667988th Torrey Pines Road
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA

```



Db 196 TCTGGACAGACTTCACTCTCACCATCAGCAGACTGGAGCCTGAAGATTTGCAGTGTA C 255

RESULT 15  
US-08-931-645-2  
; Sequence 2, Application US/08931645  
; Patent No. 6096551  
; GENERAL INFORMATION:  
; APPLICANT: Barbas, Carlos F, III  
; APPLICANT: Burton, Dennis R, III

QY 310 TACTGCCCTCGTGGGTGGT 327  
||||| | |  
Db 256 TACTGTCAGCAGTATGGT 273

Search completed: June 6, 2002, 07:38:13  
Job time: 5315 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: June 6, 2002, 07:39:08 ; Search time 2149.71 Seconds  
(without alignments)  
379.649 Million cell updates/sec

Title: US-08-881-509-5

Perfect score: 39

Sequence: 1 TGCTCGTCTTCTGTTCTGCAAGCAACTGACGTTT 39

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_on.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_to.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vl.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					

1	39	100.0	39	6	A93131	A93131 Sequence 5
2	39	100.0	1318	9	HSTCRJUNC	X98410 H. sapiens m
3	31	79.5	42	9	HS235208	AJ235208 Homo sapi
4	30.4	77.9	48	9	HSU27254	U27254 Human Isola
5	29.8	76.4	222	9	AF333590	AF333590 Homo sapi
6	29	74.4	51	9	AF043888	AF043888 Homo sapi
7	29	74.4	183	9	HUMTCACA	HL7656 Human T-cell
8	28.4	72.8	51	9	AF043877	AF043877 Homo sapi
9	28.4	72.8	51	9	AF043886	AF043886 Homo sapi
10	28.4	72.8	51	9	AF043890	AF043890 Homo sapi
11	28.2	72.3	39	9	HSU30390	U30448 Human Isola
12	28.2	72.3	63	9	HSTCRJ3	U30390 Human Isola
13	28.2	72.3	645	9	HUAE000662	X02886 Human gene
14	28.2	72.3	71153	9	HUMTCRACV	AE000662 Homo sapi
15	28.2	72.3	97630	2	AC023226	M94081 Human tcr-c
16	28.2	72.3	175053	2	AC023226	AC023226 Homo sapi
17	28	71.8	39	9	HSU30428	U30428 Human Isola
18	28	71.8	51	9	AF043875	AF043875 Homo sapi
19	28	71.8	51	9	AF043884	AF043884 Homo sapi
20	28	71.8	411	9	HUMTCVJCC	AF043884 Homo sapi
21	28	71.8	716	9	S63879	M97706 Human T-cell
22	27.8	71.3	432	9	HUMIGTCACA	S63879 TCR V alpha
23	27.6	70.8	36	6	A93133	M81774 Homo sapien
24	27	69.2	51	9	AF043874	A93133 Sequence 7
25	27	69.2	51	9	AF043876	AF043874 Homo sapi
26	27	69.2	51	9	AF043892	AF043876 Homo sapi
27	27	69.2	51	9	AF043894	AF043892 Homo sapi
28	27	69.2	87	9	HUMTCRACG	AF043894 Homo sapi
29	26.6	68.2	98	9	S69140	L42801 Homo sapien
30	26.2	67.2	1341	6	A93127	S69140 Homo sapien
31	25	64.1	214	9	HSU40776	U40776 Human T-cell
32	24.8	63.6	42	9	HSTCEL114	A93127 Sequence 1
33	24.6	63.1	46255	3	U20861	X81549 H. sapiens r
34	24	61.5	51	9	HSU14029	U20861 Caenorhabdi
35	24	61.5	234	9	AF041171	U14029 Human T cel
36	24	61.5	246	9	HSTCARAVB	AF041171 Homo sapi
37	23.2	59.5	136960	9	AC092595	Z29666 H. sapiens m
38	23.2	59.5	150899	2	AC018400	AC092595 Homo sapi
39	23.2	59.5	166598	2	AC106723	AC018400 Homo sapi
40	23.2	59.5	167026	9	AP002783	AC106723 Homo sapi
41	23.2	59.5	170603	2	AC069289	AP002783 Homo sapi
42	23.2	59.5	175810	2	AC068727	AC069289 Homo sapi
43	23	59.0	134640	9	AC027130	AC068727 Homo sapi
44	23	59.0	218939	2	AC087558	AC027130 Homo sapi
45	22.8	58.5	937	9	HS326131	AC087558 Mus muscu
						AJ326131 Homo sapi

## ALIGNMENTS

RESULT	1						
LOCUS		A93131	Sequence 5 from Patent EP0816496.	39 bp	DNA	linear	PAT 22-JAN-2000
DEFINITION		A93131					
ACCESSION		A93131					
VERSION		A93131.1	GI:6741520				
KEYWORDS		.					
SOURCE		unidentified.					
ORGANISM		unclassified.					
REFERENCE		1 (bases 1 to 39)					
AUTHORS		Schandel,D.J.					
TITLE		T-cells specific for kidney carcinoma					
JOURNAL		Patent: EP 0816496-A 5 07-JAN-1998;					
		BOEHRINGER MANNHEIM GMBH (DE)					
FEATURES		Location/Qualifiers					
source		1..39					
		/organism="unidentified"					
		/db_xref="taxon:32644"					
		<1..>39					
		/note="unnamed protein product"					
		/codon_start=1					
		/protein_id="CAB69531.1"					
CDS							

Thu Jun 6 10:09:09 2002

```

/db_xref="GI:6741521"
/translation="CLVLSGSGARQLTF"
BASE COUNT      5 a 12 c 8 g 14 t
ORIGIN

Query Match      100.0%; Score 39; DB 6; Length 39;
Best Local Similarity 100.0%; Pred. No. 1.7e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCCTCGCTCTTCCTGCTTCTGCAAGGCAACTGACCTTT 39
    |||||
Db 1 TGCCTCGCTCTTCCTGCTTCTGCAAGGCAACTGACCTTT 39

RESULT 2
HSTCRJUNC      1318 bp mRNA linear PRI 08-JAN-1997
LOCUS
DEFINITION
ACCESSION      H.sapiens mRNA for rearranged TCR junctional sequences.
VERSION
X98410
KEYWORDS
J gene; junction; T cell receptor; TCR junctional sequence; V gene.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1318)
AUTHORS
Jantzer,P.
TITLE
Tumor-infiltrating lymphocytes recognizing spontaneously arising
renal cell carcinomas express T cell receptors characteristic of a
secondary immune response
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 1318)
AUTHORS
Jantzer,P.
TITLE
Direct Submission
JOURNAL
Submitted (11-JUN-1996) P. Jantzer, Institute for Immunology,
Goethestr. 31, D-80336 Munich, FRG
FEATURES
source
1..1318
/organism="Homo sapiens"
/translation="patients 22 and 26"
/db_xref="taxon:9606"
/rearranged
<1..>1318
/notes="V gene/J gene junction"
/codon_start=1
/product="TCR junctional sequence"
/db_xref="GI:1770561"
/translation="CLVLSGSGARQLTFGXCCLVLSGSGARQLTFGXCCLVLSGSGARQLTFGXC
LVLSGSGARQLTFGXCCLVLSGSGARQLTFGXCCLVLSGSGARQLTFGXCCLV
LVLSGSGARQLTFGXCCLVLSGSGARQLTFGXCCLVLSGSGARQLTFGXCCLV
TGSARQLTFGXCCLVLSGSGARQLTFGXCCLVLSGSGARQLTFGXCCLV
VSGARQLTFGXCCLVLSGSGARQLTFGXCCLVLSGSGARQLTFGXCCLV
LVAGGNTPLVFGXCCLVLSGSGARQLTFGXCCLVLSGSGARQLTFGXCCLV
ATGSGARQLTFGXCCLVLSGSGARQLTFGXCCLVLSGSGARQLTFGXCCLV
CLVVLGSGQNLIFGXCCLVLPNAGNMLTFGXCCLVATNCRNSDLW"
BASE COUNT      206 a 329 c 352 g 344 t 87 others
ORIGIN

Query Match      100.0%; Score 39; DB 9; Length 1318;
Best Local Similarity 100.0%; Pred. No. 1.8e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCCTCGCTCTTCCTGCTTCTGCAAGGCAACTGACCTTT 39
    |||||
Db 85 TGCCTCGCTCTTCCTGCTTCTGCAAGGCAACTGACCTTT 123

RESULT 3
HSA235208
LOCUS
DEFINITION
ACCESSION      HSA235208
VERSION
U27254
KEYWORDS
U27254.1 GI:857594
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 48)
AUTHORS
Dave, V.P., Larche, M., Rencher, S.D., Koop, B.F. and Hurwitz, J.L.
TITLE
Restricted usage of T-cell receptor V alpha sequence and
variable-jointing pairs after normal T-cell development and bone
marrow transplantation
JOURNAL
Hum. Immunol. 37 (3), 178-184 (1993)
MEDLINE
94064390
REFERENCE
2 (bases 1 to 48)
AUTHORS
Hurwitz, J.L.
TITLE
Direct Submission
JOURNAL
Submitted (17-MAY-1995) Julia L. Hurwitz, Immunology, St. Jude
Children's Research Hospital, 332 N. Lauderdale, Memphis, TN 38101,
USA
FEATURES
source
Location/Qualifiers
V_region 6 a 11 g 16 t
BASE COUNT      6 a 11 g 16 t
ORIGIN

Query Match      79.5%; Score 31; DB 9; Length 42;
Best Local Similarity 87.2%; Pred. No. 0.0063;
Matches 34; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TGCCTCGCTCTTCCTGCTTCTGCAAGGCAACTGACCTTT 39
    |||||
Db 1 TGCCTCGCTCTTCCTGCTTCTGCAAGGCAACTGACCTTT 39

RESULT 4
HSA27254
LOCUS
DEFINITION
ACCESSION      HSA27254
VERSION
U27254.1 GI:857594
KEYWORDS
U27254.1 GI:857594
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 48)
AUTHORS
Dave, V.P., Larche, M., Rencher, S.D., Koop, B.F. and Hurwitz, J.L.
TITLE
Restricted usage of T-cell receptor V alpha sequence and
variable-jointing pairs after normal T-cell development and bone
marrow transplantation
JOURNAL
Hum. Immunol. 37 (3), 178-184 (1993)
MEDLINE
94064390
REFERENCE
2 (bases 1 to 48)
AUTHORS
Hurwitz, J.L.
TITLE
Direct Submission
JOURNAL
Submitted (17-MAY-1995) Julia L. Hurwitz, Immunology, St. Jude
Children's Research Hospital, 332 N. Lauderdale, Memphis, TN 38101,
USA
FEATURES
source
Location/Qualifiers
V_region 6 a 11 g 16 t
BASE COUNT      6 a 11 g 16 t
ORIGIN

Query Match      79.5%; Score 31; DB 9; Length 42;
Best Local Similarity 87.2%; Pred. No. 0.0063;
Matches 34; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TGCCTCGCTCTTCCTGCTTCTGCAAGGCAACTGACCTTT 39
    |||||
Db 1 TGCCTCGCTCTTCCTGCTTCTGCAAGGCAACTGACCTTT 39

RESULT 4
HSA27254
LOCUS
DEFINITION
ACCESSION      HSA27254
VERSION
U27254.1 GI:857594
KEYWORDS
U27254.1 GI:857594
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 48)
AUTHORS
Dave, V.P., Larche, M., Rencher, S.D., Koop, B.F. and Hurwitz, J.L.
TITLE
Restricted usage of T-cell receptor V alpha sequence and
variable-jointing pairs after normal T-cell development and bone
marrow transplantation
JOURNAL
Hum. Immunol. 37 (3), 178-184 (1993)
MEDLINE
94064390
REFERENCE
2 (bases 1 to 48)
AUTHORS
Hurwitz, J.L.
TITLE
Direct Submission
JOURNAL
Submitted (17-MAY-1995) Julia L. Hurwitz, Immunology, St. Jude
Children's Research Hospital, 332 N. Lauderdale, Memphis, TN 38101,
USA
FEATURES
source
Location/Qualifiers
V_region 6 a 11 g 16 t
BASE COUNT      6 a 11 g 16 t
ORIGIN

```

```

source
1. .48
/organism="Homo sapiens"
/isolate="M30"
/db_xref="taxon:9606"
/tissue_type="blood"
<1. .>48
/notes="encodes V alpha 5/J alpha 22 junction"
/pseudo
/evidence-experimental
/product="T-cell receptor"
7 a 15 c 10 g 16 t

CDS
BASE COUNT
ORIGIN
7 a 15 c 10 g 16 t

Query Match
Best Local Similarity 77.9%; Score 30.4; DB 9; Length 48;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 8 TCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
|||||
Db 17 TCCTATCTGGTTCTGCAAGGCAACTGACCTTT 48
|||||

RESULT 5
AF333590
LOCUS
DEFINITION
Homo sapiens clone 152.5 T cell receptor alpha chain mRNA, partial cds.
ACCESSION
AF333590
VERSION
AF333590.1 GI:13492195
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS
Talken,B.L., Bailey,C.W., Reardon,S.L., Caldwell,C.W. and
Hoffman,R.W.
TITLE
Structural analysis of TCRalpha and beta chains from human T-cell
clones specific for small nuclear ribonucleoprotein polypeptides
Sm-D, Sm-B and U1-70 kDa: TCR complementarity determining region 3
usage appears highly conserved
Scand. J. Immunol. 54 (1-2), 204-210 (2001)
21332912
11439168
PUBMED
REFERENCE
1 (bases 1 to 222)
Talken,B.L., Bailey,C.W., Schafermeyer,K.R. and Hoffman,R.W.
T cell receptor alpha and beta chain usage among anti-srRNP
reactive human T cell clones and lines
unpublished
JOURNAL
3 (bases 1 to 222)
Talken,B.L., Bailey,C.W., Schafermeyer,K.R. and Hoffman,R.W.
Direct Submission
TITLE
Submitted (05-JAN-2001) Internal Medicine, University of Missouri,
One Hospital Drive, Columbia, MO 65212, USA
JOURNAL
FEATURES
source
1. .222
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q11-q12"
/clone="152.5"
/notes="isolated from connective tissue disease patient"
<1. .>222
/codon_start=1
/product="T cell receptor alpha chain"
/protein_id="AAK28123.1"
/db_xref="GI:13492196"
/translation="SAMVFCATRGAFSGSARQLTFGSGTQLTVLPDIONPPAVYQLR
DSKSSDKSVCLTFDFDSQTNVSQKNSDY"
<1. .21
/notes="AV6S1A2N2"

V_segment
7

```

---

```

J_segment
23. .93
/notes="AJ22"
94. .>222
/notes="AC"
58 a 51 c 47 g 66 t

BASE COUNT
ORIGIN
58 a 51 c 47 g 66 t

Query Match
Best Local Similarity 76.4%; Score 29.8; DB 9; Length 222;
Matches 31; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 7 GTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
|||||
Db 31 GCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 63
|||||

RESULT 6
AF043888
LOCUS
DEFINITION
Homo sapiens patient CS-1 clone AV228 T cell receptor alpha chain
CDR3 (TCRA) mRNA, partial cds.
ACCESSION
AF043888
VERSION
AF043888.1 GI:3859395
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS
Striebach,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
Selective accumulation of related CD4+ T cell clones in the
-synovial fluid of patients with rheumatoid arthritis
J. Immunol. 161 (8), 4428-4436 (1998)
98451502
2 (bases 1 to 51)
Striebach,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
Direct Submission
TITLE
Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and
Research Center, 1400 Jackson St., Denver, CO 80206, USA
Location/Qualifiers
1. .51
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="AV228"
/tissue_type="synovial fluid"
/notes="from rheumatoid arthritis patient CS-1 [TCRAVIS3]"
[TCRAJ22]
<1. .>51
/gene="TCRA"
CDS
<1. .>51
/gene="TCRA"
/notes="Contains the 3' end of the TCRAV/TCRBV, the nDn/n,
and the 5' end of the TCRAJ/TCRBJ."
/codon_start=1
/product="T cell receptor alpha chain CDR3"
/protein_id="AAC72883.1"
/db_xref="GI:3859396"
/translation="YFCAVRISGSARQLTFG"
9 a 10 c 14 g 18 t

BASE COUNT
ORIGIN
9 a 10 c 14 g 18 t

Query Match
Best Local Similarity 74.4%; Score 29; DB 9; Length 51;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 TTCTTGTTCTGCAAGGCAACTGACCTTT 39
|||||
Db 20 TTCTTGTTCTGCAAGGCAACTGACCTTT 48
|||||

RESULT 7

```

Thu Jun 6 10:09:09 2002

```

JOURNAL      HUMTCAZA
MEDLINE      183 bp      mRNA      linear      PRI 13-JAN-1995
REFERENCE    Human T-cell receptor active alpha-chain V-region (V-J-C) mRNA,
AUTHORS      partial cds, clone AA27.
TITLE        M17666 J02992
JOURNAL      M17666.1 GI:338805
KEYWORDS     C-region; J-region; T-cell receptor; V-region; antigen receptor;
SOURCE       processed gene.
ORGANISM     Human peripheral blood lymphocyte, cDNA to mRNA, clone AA27.
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE        1 (bases 1 to 183)
JOURNAL      Klein,M.H., Concannon,P., Everett,M., Kim,L.D., Hunkapiller,T. and
AUTHORS      Hood,L.
TITLE        Diversity and structure of human T-cell receptor alpha-chain
JOURNAL      variable region genes
MEDLINE      Proc. Natl. Acad. Sci. U.S.A. 84 (19), 6884-6888 (1987)
FEATURES     88016194
SOURCE       Location/Qualifiers
            1..183
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /map="14q11.2"
            1..183
            /gene="TCRA"
            <1..>183
            /note="T-cell receptor alpha-chain V-region (V-J-C)"
            /codon_start=3
            /protein_id="AAA0645.1"
            /db_xref="GI:553669"
            /db_xref="GDB:G00-120-404"
            /translation="VKINGFEAEFKKSETSHLTPKSAHMSDAAEYFCAVTFSGSAR
            OLTFGSGTQLVLPDI"
            <1..>183
            /gene="TCRA"
            /product="T-cell receptor alpha-chain"
            114..115
            /gene="TCRA"
            /organism="Homo sapiens"
            177..178
            /gene="TCRA"
            /organism="Homo sapiens"
            44 a 43 c 43 g 53 t
            BASE COUNT      44 a 43 c 43 g 53 t
            ORIGIN          Chromosome 14q11.2.

Query Match      74.4%; Score 29; DB 9; Length 183;
Best Local Similarity 100.0%; Pred. No. 0.051;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 TTTCTGTTCTGCAAGCAACTGACCTTT 39
      |||||
Db 118 TTTCTGTTCTGCAAGCAACTGACCTTT 146

RESULT 8
AF043877 51 bp      mRNA      linear      PRI 11-NOV-1998
LOCUS     Homo sapiens patient CS-1 clone AV4 T cell receptor alpha chain
DEFINITION CDR3 (TCRA) mRNA, partial cds.
ACCESSION  AF043877
VERSION     AF043877.1 GI:3859383
KEYWORDS   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 51)
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="AV314"
            /tissue_type="synovial fluid"
            /note="from rheumatoid arthritis patient CS-1 [TCRAV16S1]
            [TCRAJ22]"
            <1..>51
            /gene="TCRA"
            <1..>51

REFERENCE    Striebach,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
AUTHORS      Selective accumulation of related CD4+ T cell clones in the
TITLE        synovial fluid of patients with rheumatoid arthritis
JOURNAL      J. Immunol. 161 (8), 4428-4436 (1998)
MEDLINE      98451502
REFERENCE    2 (bases 1 to 51)
AUTHORS      Striebach,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
TITLE        Direct Submission
JOURNAL      Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and
            Research Center, 1400 Jackson St., Denver, CO 80206, USA
FEATURES     Location/Qualifiers
            1..51
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="AV314"
            /tissue_type="synovial fluid"
            /note="from rheumatoid arthritis patient CS-1 [TCRAV16S1]
            [TCRAJ22]"
            <1..>51
            /gene="TCRA"
            <1..>51

```

```

JOURNAL      J. Immunol. 161 (8), 4428-4436 (1998)
MEDLINE      98451502
REFERENCE    2 (bases 1 to 51)
AUTHORS      Striebach,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
TITLE        Direct Submission
JOURNAL      Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and
            Research Center, 1400 Jackson St., Denver, CO 80206, USA
FEATURES     Location/Qualifiers
            1..51
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="AV4"
            /tissue_type="synovial fluid"
            /note="from rheumatoid arthritis patient CS-1 [TCRAV16S3]
            [TCRAJ22]"
            <1..>51
            /gene="TCRA"
            <1..>51
            /note="Contains the 3' end of the TCRAV, the n, and the 5'
            end of the TCRAJ."
            /codon_start=1
            /product="T cell receptor alpha chain CDR3"
            /protein_id="AAC72677.1"
            /db_xref="GI:3859384"
            /translation="YFCVVRPSGSARQLTFG"
            9 a 12 c 13 g 17 t
            BASE COUNT      9 a 12 c 13 g 17 t
            ORIGIN          72.8%; Score 28.4; DB 9; Length 51;
            Query Match      96.7%; Pred. No. 0.091;
            Best Local Similarity 96.7%; Pred. No. 0.091;
            Matches 29; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 CTTTCTGTTCTGCAAGCAACTGACCTTT 39
      |||||
Db 19 CTTTCTGTTCTGCAAGCAACTGACCTTT 48

RESULT 9
AF043886 51 bp      mRNA      linear      PRI 11-NOV-1998
LOCUS     Homo sapiens patient CS-1 clone AV314 T cell receptor alpha chain
DEFINITION CDR3 (TCRA) mRNA, partial cds.
ACCESSION  AF043886
VERSION     AF043886.1 GI:3859391
KEYWORDS   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 51)
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="AV314"
            /tissue_type="synovial fluid"
            /note="from rheumatoid arthritis patient CS-1 [TCRAV16S1]
            [TCRAJ22]"
            <1..>51
            /gene="TCRA"
            <1..>51

```

```

/gene="TCRA"
/note="Contains the 3' end of the TCRAV/TCRBV, the nDn/n,
and the 5' end of the TCRAJ/TCRBJ."
/codon_start=1
/product="T cell receptor alpha chain CDR3"
/protein_id="AAC72681.1"
/db_xref="GI:3859392"
/translation="YFCVRLTGSARQLTFG"
10 a 11 c 13 g 17 t
BASE COUNT
ORIGIN

Query Match 72.8%; Score 28.4; DB 9; Length 51;
Best Local Similarity 96.7%; Pred. No. 0.091;
Matches 29; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 CTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
|||||
Db 19 CTTACTGGTTCTGCAAGGCAACTGACCTTT 48
|||||

RESULT 10
AF043890
LOCUS
DEFINITION
Homo sapiens patient CS-1 clone AV109 T cell receptor alpha chain
CDR3 (TCRA) mRNA, partial cds.
ACCESSION
AF043890
VERSION
AF043890.1 GI:3859399
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 51)
AUTHORS
Striebach,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
TITLE
Selective accumulation of related CD4+ T cell clones in the
synovial fluid of patients with rheumatoid arthritis
J. Immunol. 161 (8), 4428-4436 (1998)
98451502
MEDLINE
2 (bases 1 to 51)
AUTHORS
Striebach,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
DIRECT SUBMISSION
TITLE
Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and
Research Center, 1400 Jackson St., Denver, CO 80206, USA
JOURNAL
Location/Qualifiers
1..51
FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="AV109"
/tissue_type="synovial fluid"
/note="from rheumatoid arthritis patient CS-1 [TCRAVIS3]
[TCRAJ22]"
<1..>51
<1..>51
<1..>51
<1..>51
/gene="TCRA"
/gene="TCRA"
/note="Contains the 3' end of the TCRAV/TCRBV, the nDn/n,
and the 5' end of the TCRAJ/TCRBJ."
/codon_start=1
/product="T cell receptor alpha chain CDR3"
/protein_id="AAC72685.1"
/db_xref="GI:3859400"
/translation="YFCVRRSGSARQLTFG"
8 a 11 c 15 g 17 t
BASE COUNT
ORIGIN

Query Match 72.8%; Score 28.4; DB 9; Length 51;
Best Local Similarity 96.7%; Pred. No. 0.091;
Matches 29; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 CTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
|||||

```

```

Db 19 CGTTCTGGTTCTGCAAGGCAACTGACCTTT 48
|||||
RESULT 11
HSU30448
LOCUS
DEFINITION
Human isolate M94 T-cell receptor alpha V-J junction (TCR Valpha
3/J alpha 22) mRNA, partial cds.
ACCESSION
U30448
VERSION
U30448.1 GI:915515
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 39)
AUTHORS
Dave,V.P., Larche,M., Rencher,S.D., Koop,B.F. and Hurwitz,J.L.
TITLE
Restricted usage of T-cell receptor V alpha sequence and
variable-joining pairs after normal T-cell development and bone
marrow transplantation
Hum. Immunol. 37 (3), 178-184 (1993)
94064390
MEDLINE
2 (bases 1 to 39)
AUTHORS
Hurwitz,J.L.
DIRECT SUBMISSION
TITLE
Submitted (22-JUN-1995) Julia L. Hurwitz, St. Jude Children's
Research Hospital, Immunology, 332 N. Lauderdale, Memphis, TN
38101, USA
JOURNAL
Location/Qualifiers
1..39
FEATURES
source
/organism="Homo sapiens"
/isolate="M94"
/db_xref="taxon:9606"
/tissue_type="blood"
1..39
/gene="TCR Valpha 3/J alpha 22"
<1..>39
/gene="TCR Valpha 3/J alpha 22"
/codon_start=1
/product="T-cell receptor alpha V-J junction"
/protein_id="AA73621.1"
/db_xref="GI:915516"
/translation="FCVSSGSARQLTF"
5 a 11 c 8 g 15 t
BASE COUNT
ORIGIN

Query Match 72.3%; Score 28.2; DB 9; Length 39;
Best Local Similarity 90.9%; Pred. No. 0.11;
Matches 30; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 7 GTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39
|||||
Db 7 GTCCTCTCTGGTTCTGCAAGGCAACTGACCTTT 39
|||||

RESULT 12
HSU30390
LOCUS
DEFINITION
Human isolate M101 T-cell receptor alpha V-J junction (TCR Valpha
5/J alpha 22) mRNA, partial cds.
ACCESSION
U30390
VERSION
U30390.1 GI:915415
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 63)
AUTHORS
Dave,V.P., Larche,M., Rencher,S.D., Koop,B.F. and Hurwitz,J.L.
TITLE
Restricted usage of T-cell receptor V alpha sequence and
variable-joining pairs after normal T-cell development and bone
marrow transplantation

```

JOURNAL Hum. Immunol. 37 (3), 178-184 (1993)  
 MEDLINE 94064390  
 REFERENCE 2 (bases 1 to 63)  
 AUTHORS Hurwitz, J.L.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-JUN-1995) Julia L. Hurwitz, St. Jude Children's Research Hospital, Immunology, 332 N. Lauderdale, Memphis, TN 38101, USA

FEATURES  
 source Location/Qualifiers  
 1..63  
 /organism="Homo sapiens"  
 /isolate="M101"  
 /db\_xref="taxon:9606"  
 /tissue\_type="blood"  
 gene 1..63  
 /gene="TCR Valpha 5/J alpha 22"  
 CDS <1..>63  
 /gene="TCR Valpha 5/J alpha 22"  
 /codon\_start=1  
 /product="T-cell receptor alpha V-J junction"  
 /protein\_id="AA037579.1"  
 /db\_xref="GI:915416"  
 /translation="TYLCALDPRVTSSGSARQLTF"  
 BASE COUNT 11 a 20 c 12 g 20 t  
 ORIGIN

Query Match 72.3%; Score 28.2; DB 9; Length 63;  
 Best Local Similarity 90.9%; Pred. No. 0.11;  
 Matches 30; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 GTCCTTTCTGTTCTGCAAGCACTGACCTTT 39  
 || |||||  
 Db 31 GTTCTTCTGTTCTGCAAGCACTGACCTTT 63

RESULT 13  
 HSTCRAJ3  
 LOCUS Human gene for T-cell receptor alpha chain J region. PRI 11-APR-1995  
 DEFINITION 645 bp DNA linear  
 ACCESSION X02886  
 VERSION X02886.1 GI:36928  
 KEYWORDS joining region; T-cell receptor; T-cell receptor alpha.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 645)  
 AUTHORS Yoshikai, Y., Clark, S.P., Taylor, S., Sohn, U., Wilson, B.I., Minden, M.D., and Mak, T.W.  
 TITLE Organization and sequences of the variable, joining and constant region genes of the human T-cell receptor alpha-chain  
 JOURNAL Nature 316 (6031), 837-840 (1985)  
 MEDLINE 85296333  
 FEATURES  
 source Location/Qualifiers  
 1..645  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 misc\_feature 377..385  
 /note="nonamer recombination signal"  
 misc\_feature 398..404  
 /note="heptamer recombination signal"  
 J\_segment 405..467  
 /note="J segment (405 is 2nd base in codon) (467 is 1st base in codon)"  
 BASE COUNT 201 a 107 c 149 g 188 t  
 ORIGIN

Query Match 72.3%; Score 28.2; DB 9; Length 645;  
 Best Local Similarity 90.9%; Pred. No. 0.12;  
 Matches 30; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 GTCCTTTCTGTTCTGCAAGCACTGACCTTT 39  
 || |||||  
 Db 404 GTTCTTCTGTTCTGCAAGCACTGACCTTT 436

RESULT 14  
 HUA000662  
 LOCUS Homo sapiens T-cell receptor alpha delta locus from bases 1000498 to 1071650 (section 5 of 5) of the Complete Nucleotide Sequence.  
 DEFINITION 71153 bp DNA linear PRI 15-JUN-2001  
 ACCESSION HUA000662  
 VERSION AE000662  
 KEYWORDS AE000662.1 GI:2358068  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 71153)  
 AUTHORS Koop, B.F., Rowen, L., Wang, K., Kuo, C.L., Seto, D., Lenstra, J.A., Howard, S., Shan, W., Deshpande, P., and Hood, L.  
 TITLE The human T-cell receptor TCRAC/TCRDC (C alpha/C delta) region: organization, sequence, and evolution of 97.6 Kb of DNA  
 JOURNAL Genomics 19 (3), 478-493 (1994)  
 MEDLINE 94245236  
 REMARK This citation covers from bases 966383-1064019  
 REFERENCE 2 (bases 1 to 71153)  
 AUTHORS Boyesen, C., Simon, M.I., and Hood, L.  
 TITLE Analysis of the 1.1-Mb human alpha/delta T-cell receptor locus with bacterial artificial chromosome clones  
 JOURNAL Genome Res. 7 (4), 330-338 (1997)  
 MEDLINE 97264339  
 PUBMED 9110172  
 REFERENCE 3 (bases 1 to 71153)  
 AUTHORS Boyesen, C., Inyoub, L., Smith, T.M., Smit, A., Wang, K., Rowen, L., and Hood, L.  
 TITLE T-Cell Receptor Alpha Delta Locus Complete Nucleotide Sequence  
 JOURNAL Unpublished  
 REMARK This citation covers bases 1-983545 and bases 1064020-1071650  
 REFERENCE 4 (bases 1 to 71153)  
 AUTHORS Boyesen, C., Lee, I., Smith, T.M., Smit, A., Kai, W., Lee, R., and Leroy, H.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-JUL-1997) Department of Molecular Biotechnology, University of Washington, Box 357730, Seattle, Washington 98195, USA  
 REMARK Complete nucleotide sequence of the human T-cell receptor alpha delta locus  
 COMMENT Sequencing method: high redundancy shotgun. Interspersed Repeats were identified with RepeatMasker (available from <http://ftp.genome.washington.edu/RM/RepeatMasker.html>) Simple sequence repeats were identified with sputnik (available from <http://serac.mbt.washington.edu/~chrissa/software/sputnik.html>).  
 FEATURES  
 source Location/Qualifiers  
 1..71153  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 repeat\_region 2362..2660  
 /rpt\_family="AluSc"  
 repeat\_region 4274..4574  
 /rpt\_family="AluSc"  
 repeat\_region 9569..9796  
 /rpt\_family="MIR"  
 repeat\_region 10535..10834  
 /rpt\_family="AluY"  
 repeat\_region 10835..11138  
 /rpt\_family="AluY"  
 repeat\_region complement(13900..14190)  
 /rpt\_family="AluSc"  
 repeat\_region complement(20506..20810)  
 /rpt\_family="AluY"  
 repeat\_region 20928..21223  
 /rpt\_family="AluSc"  
 repeat\_region 23083..23335  
 /rpt\_family="MIR"



```

repeat_region 25320..25618
                /rpt_family="AluSq"
repeat_region complement(27167..27240)
                /rpt_family="MIR"
repeat_region 27306..27428
                /rpt_family="MIR2"
repeat_region complement(30158..30403)
                /rpt_family="MIR"
repeat_region 33727..33878
                /rpt_family="MIR"
repeat_region complement(33920..34213)
                /rpt_family="AluYa5"
repeat_region complement(35275..35553)
                /rpt_family="MIR2"
repeat_region complement(36848..36886)
                /rpt_family="L1MA3"
repeat_region 39321..39398
                /rpt_family="MIR"
repeat_region 40206..40261
                /rpt_family="MIR"
repeat_region complement(40433..40659)
                /rpt_family="MIR"
repeat_region complement(40883..40848)
                /rpt_family="AluXs"
repeat_region complement(41322..41444)
                /rpt_family="MIR"
repeat_region complement(45732..45964)
                /rpt_family="MIR"
repeat_region 46927..46992
                /rpt_family="MIR2"
repeat_region complement(47281..47388)
                /rpt_family="MIR2"
repeat_region 49066..49073
                /rpt_type=tandem
repeat_region /rpt_unit=tat
                49734..50032
repeat_region /rpt_family="AluY"
                51737..51940
repeat_region /rpt_family="MER3"
                complement(52892..53047)
repeat_region 54558..54732
                /rpt_family="MIR"
repeat_region /rpt_family="MIR2"
                complement(54749..55052)
repeat_region 55059..55302
                /rpt_family="AluSq"
repeat_region 57276..57570
                /rpt_family="MIR2"
repeat_region 59210..59460
                /rpt_family="AluJo"
repeat_region /rpt_family="MER46"
                complement(61706..61769)
repeat_region 63506..63580
                /rpt_family="MIR2"
repeat_region 63695..63862
                /rpt_family="MIR"
repeat_region 63931..63951
                /rpt_family="L1ME3"
repeat_region /rpt_type=tandem
                /rpt_unit=AC
repeat_region 64131..64281
                /rpt_family="MIR"
repeat_region 65845..66137
                /rpt_family="AluXs"
repeat_region complement(66237..66480)
                /rpt_family="MIR2"
repeat_region complement(68187..68457)
                /rpt_family="AluJo"
repeat_region complement(68480..68777)
                /rpt_family="AluSq"
repeat_region complement(68778..68837)
                /rpt_family="FAM"
repeat_region 69326..69628

```

```

repeat_region /rpt_family="AluSp"
                complement(69688..69766)
repeat_region /rpt_family="MER58A"
                69845..69897
repeat_region /rpt_family="MER5A"
                complement(69907..70048)
repeat_region /rpt_family="MER5A"
                complement(70060..70225)
                /rpt_family="MER5B"
BASE COUNT 20125 a 15457 c 16410 g 19161 t
ORIGIN

```

```

Query Match 72.3%; Score 28.2; DB 9; Length 71153;
Best Local Similarity 90.9%; Pred. No. 0.13;
Matches 30; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 GTCCTTTCTGTTCTGCAAGGCAACTGACCTTT 39
    || ||||| ||||| ||||| ||||| |||||
Db 28162 GTTCTCTGTTCTGCAAGGCAACTGACCTTT 28194

```

## RESULT 15

```

HUMTCRADCV
LOCUS Human Tcr-C-delta gene, exons 1-4; Tcr-V-delta gene, exons 1-2;
DEFINITION T-cell receptor alpha (Tcr-alpha) gene, J1-J61 segments; and
                Tcr-C-alpha gene, exons 1-4.
ACCESSION M94081
VERSION M94081.1 GI:2627263
KEYWORDS T-cell receptor C-alpha; T-cell receptor C-delta; T-cell receptor
                V-delta; T-cell receptor alpha.
SOURCE human.
ORGANISM Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 97630)
AUTHORS Koop,B.F., Rowen,L., Wang,K., Kuo,C.L., Seto,D., Lenstra,J.A.,
                Howard,S., Shan,W., Deshpande,P. and Hood,L.
TITLE The human T-cell receptor TCRAC/TCRDC (C alpha/C delta) region:
                organization, sequence, and evolution of 97.6 kb of DNA
JOURNAL Genomics 19 (3), 478-493 (1994)
MEDLINE 94245236
PUBMED 8188290
COMMENT On Nov 20, 1997 this sequence version replaced gi:339242.
FEATURES
                Location/Qualifiers
                1..97630
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /germline
                3168..3464
                /gene="Tcr-C-delta"
                /number=1
                3168..6809
                /gene="Tcr-C-delta"
                4028..4098
                /gene="Tcr-C-delta"
                /number=2
                4458..4572
                /gene="Tcr-C-delta"
                /number=3
                5787..6809
                /gene="Tcr-C-delta"
                /number=4
                9295..9868
                /gene="Tcr-V-delta"
                9295..9598
                /gene="Tcr-V-delta"
                /number=1
                9832..9868
                /gene="Tcr-V-delta"
                /number=2
                <15555..>85296

```

```

                exon
                3168..3464
                /gene="Tcr-C-delta"
                /number=1
                gene
                3168..6809
                /gene="Tcr-C-delta"
                exon
                4028..4098
                /gene="Tcr-C-delta"
                /number=2
                exon
                4458..4572
                /gene="Tcr-C-delta"
                /number=3
                exon
                5787..6809
                /gene="Tcr-C-delta"
                /number=4
                gene
                9295..9868
                /gene="Tcr-V-delta"
                exon
                9295..9598
                /gene="Tcr-V-delta"
                /number=1
                exon
                9832..9868
                /gene="Tcr-V-delta"
                /number=2
                gene
                <15555..>85296

```



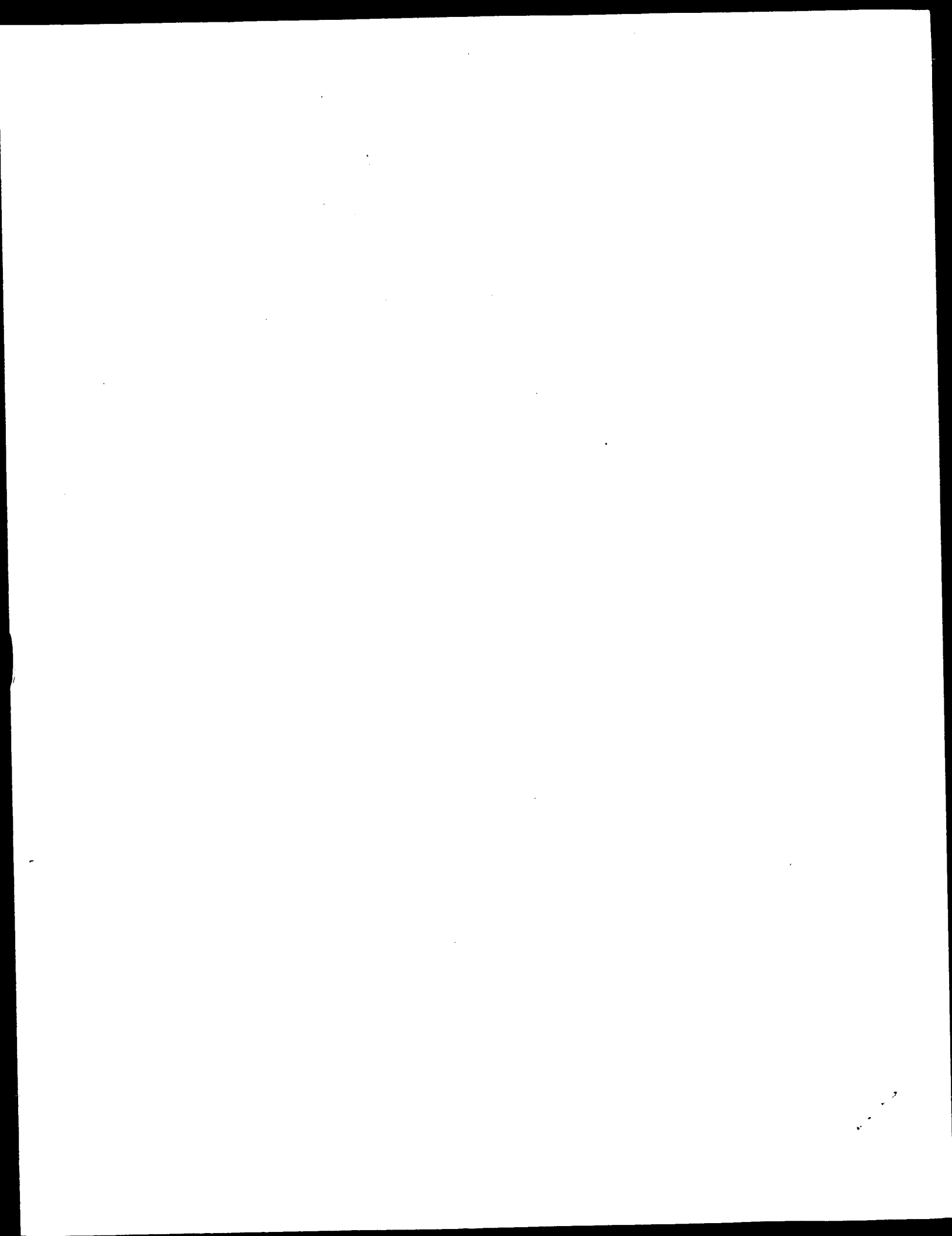
CDS

```
<37899..>37961
/gene="Tcr-alpha"
/standard_name="possible J 41 gene segment"
/note="putative"
/codon_start=1
```

```
Query Match          72.3%  Score 28.2; DB 9; Length 97630;
Best Local Similarity 90.9%; Pred. No. 0.13;
Matches 30; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 7 GTCCTTTCTGGTCTGCAAGGCAACTGACCTTT 39
    || ||||| ||||| ||||| ||||| |||||
Db 62271 GTTCTTCTGGTCTGCAAGGCAACTGACCTTT 62303
```

Search completed: June 6, 2002, 07:39:13  
Job time: 5376 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 06:54:26 ; Search time 2649.09 seconds  
(without alignments)  
198.703 Million cell updates/sec

Title: US-08-881-509-5  
Perfect score: 39  
Sequence: 1 TGCCTGCTCTTCTGCTTCGCAAGCAACTGACCTTT 39

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estinu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	30.4	77.9	564	12	AQ221133	AQ221133 HS_3223_B
2	26.2	67.2	127	10	BJ140929	BJ140929 BJ140929
3	24.6	63.1	609	12	AZ570267	AZ570267 272PVD08
4	24.6	63.1	648	10	BJ138927	BJ138927 BJ138927
5	23	59.0	292	9	BB231582	BB231582 BB231582
6	22.8	58.5	294	9	AA234451	AA234451 zr74a08.r
7	22.2	56.9	316	9	BB253473	BB253473 BB253473
8	22.2	56.9	393	10	BI289485	BI289485 BI289485
9	22	56.4	157	10	BG993065	BG993065 RC4-HT110
10	22	56.4	267	10	BG375218	BG375218 UT-R-CV1-f
11	22	56.4	315	9	AA955644	AA955644 BI294666
12	22	56.4	336	10	BG375188	BG375188 BI293376
13	22	56.4	420	9	AI176100	AI176100 EST219678
14	22	56.4	470	10	BI294666	BI294666 BI293376
15	22	56.4	508	10	BI293376	BI293376 BI293376
16	22	56.4	564	12	AQ683171	AQ683171 HS_5378_B
17	22	56.4	633	9	AI233243	AI233243 EST229931

18	22	56.4	899	12	CNS0407S	AL269561 Tetraodon
19	22	56.4	1101	12	CNS05SV2	AL352343 Tetraodon
20	21.6	55.4	470	12	AQ993790	AQ993790 RPCI-23-3
21	21.6	55.4	551	12	AQ491072	AQ491072 RPCI-11-2
22	21.6	55.4	619	12	BH062049	BH062049 RPCI-24-3
23	21.6	55.4	664	12	AZ313505	AZ313505 1M0029107
24	21.6	55.4	667	9	B8639586	B8639586 B8639586
25	21.6	55.4	796	10	BF120330	BF120330 601756068
26	21.4	54.9	269	9	AV294476	AV294476 AV294476
27	21.4	54.9	275	9	BB367011	BB367011 BB367011
28	21.4	54.9	279	9	BB360913	BB360913 BB360913
29	21.4	54.9	282	9	BB461910	BB461910 BB461910
30	21.4	54.9	285	9	BB345796	BB345796 BB345796
31	21.4	54.9	291	9	BB079644	BB079644 BB079644
32	21.4	54.9	294	9	BB402658	BB402658 BB402658
33	21.4	54.9	294	9	BB455396	BB455396 BB455396
34	21.4	54.9	296	9	BB533795	BB533795 BB533795
35	21.4	54.9	305	9	BB375951	BB375951 BB375951
36	21.4	54.9	317	9	BB097763	BB097763 BB097763
37	21.4	54.9	358	9	BB813655	BB813655 BB813655
38	21.4	54.9	370	9	BB813418	BB813418 BB813418
39	21.4	54.9	418	9	BB711165	BB711165 BB711165
40	21.4	54.9	426	9	BB820622	BB820622 BB820622
41	21.4	54.9	432	9	BB87881	BB87881 BB87881
42	21.4	54.9	432	9	BB824353	BB824353 BB824353
43	21.4	54.9	449	9	BB820102	BB820102 BB820102
44	21.4	54.9	470	9	BB822403	BB822403 BB822403
45	21.4	54.9	477	9	AI150530	AI150530 qf41f05.x

## ALIGNMENTS

### RESULT 1

AQ221133  
LOCUS HS\_3223\_B2\_E10\_T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3223 Col=20 Row=J, DNA sequence.  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 564)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE  
COMMENT

Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 3223 row: J column: 20  
Class: BAC ends  
High quality sequence stop: 564.  
Location/Qualifiers  
1..564  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_plate="3223 Col=20 Row=J"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"



Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshini@genes.nig.ac.jp.

## FEATURES

source  
 1. .648  
 /organism="Caenorhabditis elegans"  
 /strain="N2"  
 /db\_xref="taxon:6239"  
 /clone\_lib="unpublished oligo-capped cDNA library, C."  
 /elegant LI stage"  
 /sex="hermaphrodite"  
 /tissue\_type="whole animal"  
 /dev\_stage="L1"  
 202 a 107 c 145 g 194 t

BASE COUNT  
 ORIGIN

Query Match 63.1%; Score 24.6; DB 10; Length 648;  
 Best Local Similarity 76.9%; Pred. NO. 31;  
 Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 TGCTCGTCTCTGTTCTGCAAGGCAACTGACCTTT 39  
 ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 639 TGCTCGTCTCTGTTCTGTTCCAAAGGCAACTGACCTTT 601

## RESULT 5

BB231582 292 bp mRNA linear EST 04-JUL-2000  
 LOCUS BB231582 RIKEN full-length enriched, 3 days neonate thymus Mus  
 DEFINITION musculus cDNA clone A63003E12 3', mRNA sequence.  
 ACCESSION BB231582  
 VERSION BB231582.1 GI:8911437

SOURCE house mouse.  
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 292)

## REFERENCE

AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci  
 P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,  
 Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,  
 Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,  
 Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,  
 Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,  
 Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata  
 Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,  
 Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya  
 T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamazaki,I.,  
 Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino  
 M., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Mouse ESTs (Konno,H., et al.)  
 Unpublished (2000)

## TITLE

COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki  
 N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Thermotranscription and thermoactivation of thermolabile enzymes by  
 trehalose and its application for the synthesis of full length  
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
 Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,  
 Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki  
 Y. and Hayashizaki,Y.  
 Automated filtration-based high-throughput plasmid preparation  
 system. Genome Res. 9 (5), 463-470 (1999)

Carninci,P. and Hayashizaki,Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
 19-44 (1999)  
 Please visit our web site (http://genome.rtc.riken.go.jp) for  
 further details.

## FEATURES

source

1. .292  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_lib="RIKEN full-length enriched, 3 days neonate  
 thymus"  
 /tissue\_type="thymus"  
 /dev\_stage="3 days neonate"  
 /lab\_host="DH10B"  
 /note="Site\_1: SalI; Site\_2: BamHI; cDNA library was  
 prepared and sequenced in Mouse Genome Encyclopedia  
 Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in  
 RIKEN. Division of Experimental Animal Research in Riken  
 contributed to prepare mouse tissues. 1st strand cDNA was  
 primed with a primer [5'  
 GAGAGAGAGAGATCCAGAGCTTTTCTTTTNN 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse  
 transcriptase and subsequently enriched for full-length by  
 cap-trapper. cDNA went through one round of normalization  
 to Rot = 20.0 and subtraction to Rot = 459.0. Second  
 strand cDNA was prepared with the primer adapter of  
 sequence [5' GAGAGAGATTCGATTAAATTAATCCCCCCCCCCC  
 3']. cDNA was cleaved with XhoI and BamHI. Vector: a  
 modified pBluescript KS(+) after bulk excision from Lambda  
 FLC I."

BASE COUNT 69 a 91 c 52 g 80 t  
 ORIGIN

Query Match 59.0%; Score 23; DB 9; Length 292;  
 Best Local Similarity 74.4%; Pred. NO. 1e+02;  
 Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 TGCTCGTCTCTGTTCTGCAAGGCAACTGACCTTT 39  
 ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 115 TCCCCATTCATTCTGCTTCGCGAGGGGACTTGACCTTT 153

## RESULT 6

AA234451/c 294 bp mRNA linear EST 03-MAR-1997  
 LOCUS AA234451 zr74a08.r1 Soares\_NhMMPU\_S1 Homo sapiens cDNA clone IMAGE:669110 5',  
 DEFINITION similar to WP:R90.1 CE06325 PROTEIN KINASE ;, mRNA sequence.

ACCESSION AA234451  
 VERSION AA234451.1 GI:1858970

## KEYWORDS

SOURCE EST.

ORGANISM human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS

1 (bases 1 to 294)  
 Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman  
 M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,  
 Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston  
 R., Williamson,A., Wohldmann,P. and Wilson,R.  
 The WashU-Merck EST Project  
 Unpublished (1995)

## TITLE

JOURNAL

COMMENT

Contact: Wilton RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 282.

# FEATURES

Location/Qualifiers  
1. .294  
/organism="Homo sapiens"  
/db\_xref="GDB:5563077"  
/db\_xref="taxon:9606"  
/clone="INAGE:669110"  
/clone\_lib="Soares\_NhMPu\_S1"  
/tissue\_type="Pooled human melanocyte, fetal heart, and pregnant uterus"  
/lab\_host="DH10B"  
/note="Organ: mixed (see below); Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NBHM, pregnant uterus NBHPU, and fetal heart NBH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."  
76 a 59 c 79 g 80 t

# BASE COUNT

ORIGIN  
Query Match 58.5%; Score 22.8; DB 9; Length 294;  
Best Local Similarity 79.4%; Pred. No. 1.3e+02;  
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 CCTCGTCTTCTGTTCTGCAAGCAACTGACC 36  
||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||  
Db 283 CCTTGCCTTTATTTTCTCCAGGCGAGTGACC 250

# RESULT

BB253473 316 bp mRNA linear EST 06-JUL-2000  
LOCUS  
DEFINITION  
BB253473 RIKEN full-length enriched, 7 days neonate cerebellum Mus musculus cDNA clone A730057K23 3', similar to L03305 Mus musculus core-binding factor, mRNA sequence.

BB253473  
BB253473.1 GI:8946219

# ACCESSION

VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus  
house mouse  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

# REFERENCE

AUTHORS  
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigenoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yananaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Konno, H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp/  
URL: http://genome.gsc.riken.go.jp/

# TITLE

JOURNAL  
COMMENT

# FEATURES

## SOURCE

Location/Qualifiers  
1. .316  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="A730057K23"  
/clone\_lib="RIKEN full-length enriched, 7 days neonate cerebellum"  
/tissue\_type="cerebellum"  
/dev\_stage="7 days neonate"  
/lab\_host="DH10B"  
/note="Site\_1: Salt; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGATCTCGAGTAAATTAATCCGCCGCCGCCGCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 82 a 64 c 44 g 125 t 1 others  
ORIGIN

Query Match 56.9%; Score 22.2; DB 9; Length 316;

Best Local Similarity 77.1%; Pred. No. 2.2e+02;  
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 5 TCGTCTTCTGTTCTGCAAGCAACTGACCTTT 39

Db 45 TCTTCTTACTGGTTCGCACTGCTTCTGTCGTT 79

# RESULT

8  
LOCUS  
DEFINITION  
BI289485/c 393 bp mRNA linear EST 19-JUL-2001  
UI-R-DK0-cfh-f-04-0-UI.s1 UI-R-DK0 Rattus norvegicus cDNA clone  
UI-R-DK0-cfh-f-04-0-UI 3', mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
EST.  
BI289485  
BI289485.1 GI:14947104

# ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
1 (bases 1 to 393)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477

N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.



## COMMENT

Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized rat aorta pool library CDNA Library Preparation: M.B. Soares Lab Clone Distribution: clones will be available through Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=yes.

## FEATURES

Location/Qualifiers  
1. .393  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone\_lib="UI-R-DK0"  
/clone="UI-R-DK0-cfh-f-04-0-UI"  
/dev\_stage="ADULT"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-DK0 library is a subtracted library derived from a mixture of five individually tagged normalized rat libraries: brain-nRBP (20%), heart-nRHP (20%), kidney-nRKP (20%), aorta-nRAP (20%), and placenta-nRPP (20%). Each original library was constructed from a mixture of equal amounts of RNA from seven different developmental time-points: embryonic day 17, embryonic day 19, embryonic day 21, adult day 1, adult day 12, adult day 75, and adult day 200. (Exception: the aorta pool does not contain embryonic day 17 RNA and the placenta pool contains only the three embryonic stages). Each library was normalized individually according to the procedure described by Bonaldo, Lennon & Soares (Genome Research 6: 791-806, 1996). For construction of the DK0 subtracted library, plasmid DNA from each of the five individually tagged normalized libraries was mixed in the proportions specified above and electroporated into competent bacteria for production of single-stranded circular DNA representing the pool of libraries. Single-stranded circular DNA representing these five normalized libraries was then used as a tracer in a subtractive hybridization with a driver (PCR amplified inserts from a plasmid DNA template preparation) comprising: a) a set of about 1,000 arrayed clones from each of the five non-normalized libraries of brain (CT0s), heart (CS0s), kidney (CU0s), aorta (CW0s), and placenta (CX0s). The resulting pool of approximately 5,000 clones represented about 33.3% of the final driver population. A set of about 2,000 arrayed clones from each of the five normalized libraries of brain (CT0), heart (CS0), kidney (CU0), aorta (CW0), and placenta (CX0). The resulting pool of about 10,000 clones represented about 66.6% of the final driver population.  
TAG\_LIB=UI-R-DK0  
TAG\_TISSUE=rat aorta pool  
TAG\_SEQ=CTGTAGGATC"

BASE COUNT 119 a 81 c 90 g 103 t  
ORIGIN

Query Match 56.9%; Score 22.2; DB 10; Length 393;  
Best Local Similarity 77.1%; Pred. No. 2.3e+02;  
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 4 CTCGTCTCTTCTGCTCGAAGCAACTGACCTT 38  
||||||| ||| ||||| || || |||||

Db 188 CTCGTCTTATTGGCTCTGCAGGTCTTGGCCTT 154

RESULT 9  
BG993065/c

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

157 bp mRNA linear EST 13-JUN-2001  
RC4-HT1109-060201-013-c07 HT1109 Homo sapiens cDNA, mRNA sequence.  
BG993065  
BG993065.1 GI:14397135  
EST  
human.

REFERENCE  
AUTHORS

Homo sapiens  
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 157)  
Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL  
MEDLINE  
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC4&t2=RC4-HT1109-  
060201-013-c07&t3=2001-02-06&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 25  
High quality sequence stop: 139.

FEATURES

source

Location/Qualifiers  
1. .157  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT1109"  
/dev\_stage="Adult"

/note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 43 a 37 c 45 g 32 t  
ORIGIN

Query Match 56.4%; Score 22; DB 10; Length 157;  
Best Local Similarity 73.7%; Pred. No. 2.2e+02;  
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 1 TGCCTCGCTCTTCTGCTCTGCAAGCAACTGACCTT 38  
||||| ||||| ||| ||||| ||| |||||

Db 39 TGCATCTCTCTCTCGAAGTCAAGCAACTGACCTT 2

RESULT 10  
BG375218/c

LOCUS  
DEFINITION  
ACCESSION  
VERSION

267 bp mRNA linear EST 12-MAR-2001  
BG375218  
UI-R-CV1-bsu-f-09-0-UI.s1 UI-R-CV1 Rattus norvegicus cDNA clone  
BG375218  
UI-R-CV1-bsu-f-09-0-UI 3', mRNA sequence.  
BG375218.1 GI:13299690



```

ACCESSION      BG375188
VERSION        BG375188.1  GI:13299660
KEYWORDS       EST.
SOURCE         Norway rat.
ORGANISM       Rattus norvegicus
               Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
               Rattus.
REFERENCE      1 (bases 1 to 336)
AUTHORS        Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE          Normalization and subtraction: two approaches to facilitate gene
               discovery
JOURNAL        Genome Res. 6 (9), 791-806 (1996)
MEDLINE        97044477
COMMENT        Contact: Soares, MB
               Program for Rat Gene Discovery and Mapping
               University of Iowa
               451 Eckstein Medical Research Building Iowa City, IA 52242, USA
               Tel: 319 335 8250
               Fax: 319 335 9565
               Email: msoares@blue.weeg.uiowa.edu
               The sequence contained an oligo-dt track that was present in the
               oligonucleotide that was used to prime the synthesis of first
               strand cDNA and therefore this may represent a bonafide poly A
               tail. The sequence tag present in the cDNA between the NotI site
               and the oligo-dt track served to verify it as a clone from the
               normalized rat eye library cDNA library Preparation: M.B. Soares
               Lab Clone distribution: clones will be available through Research
               Genetics (www.resgen.com)
               Seq primer: M13 Forward
               POLYA=Yes.

FEATURES             source
   source            Location/Qualifiers
       1..336
       /organism="Rattus norvegicus"
       /strain="Sprague-Dawley"
       /db_xref="taxon:10116"
       /clone="UI-R-CV1-bsu-b-09-0-UI"
       /clone_lib="UI-R-CV1"
       /dev_stage="ADULT"
       /lab_host="DH10B (Life Technologies)"
       /note="Vector: pT73B-Pac (Pharmacia) with a modified
               polylinker; Site1: Not I; Site2: Eco RI; The UI-R-CV1
               library is a normalized library constructed from rat eye
               tissue. For a detailed description of the library from
               which this clone was derived, please visit our web site at
               rategen.eng.uiowa.edu. The subtraction has been previously
               described in (Bonaldo, Lennon and Soares, Genome Research
               6:791-806, 1996)
       TAG_L1B="UI-R-CV1"
       TAG_TISSUE="rat eye"
       TAG_SEQ="CAGCC"

BASE COUNT      80 a 100 c 71 g 85 t

ORGANISM         Rattus sp.
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
               Rattus.
REFERENCE      1 (bases 1 to 420)
AUTHORS        Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
               Kerlavage,A.R. and Adams,M.D.
TITLE          Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
               Gene Index
JOURNAL        Unpublished (1998)
MEDLINE        Other_ESTS: TC48415
COMMENT        Contact: Lee, NH
               The Institute for Genomic Research
               9712, Medical Center Drive, Rockville, MD 20850, USA
               Tel: (301)-838-3529
               Fax: (301)-838-0208
               Email: nhlee@tigr.org
               Seq primer: M13-21.
               Location/Qualifiers
               1..420
               /organism="Rattus sp."
               /db_xref="ATCC (Inhost):2031158"
               /db_xref="taxon:10118"
               /clone="ROVBK59"
               /clone_lib="Normalized rat ovary, Bento Soares"
               /note="Organ: ovary; Vector: pT733Pac; Site_1: ECORI;
               Site_2: NotI"

BASE COUNT      105 a 130 c 98 g 87 t

ORGANISM         Rattus norvegicus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
               Rattus.
REFERENCE      1 (bases 1 to 470)
AUTHORS        Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE          Normalization and subtraction: two approaches to facilitate gene
               discovery
JOURNAL        Genome Res. 6 (9), 791-806 (1996)
MEDLINE        97044477
COMMENT        Contact: Soares, MB
               Program for Rat Gene Discovery and Mapping
               University of Iowa
               451 Eckstein Medical Research Building Iowa City, IA 52242, USA
               Tel: 319 335 8250
               Fax: 319 335 9565
               Email: msoares@blue.weeg.uiowa.edu
               The sequence contained an oligo-dt track that was present in the
               oligonucleotide that was used to prime the synthesis of first
               strand cDNA and therefore this may represent a bonafide poly A
               tail. The sequence tag present in the cDNA between the NotI site
               and the oligo-dt track served to identify it as a clone from the
               normalized rat brain pool library cDNA library Preparation: M.B.
               Soares Lab Clone distribution: clones will be available through

```



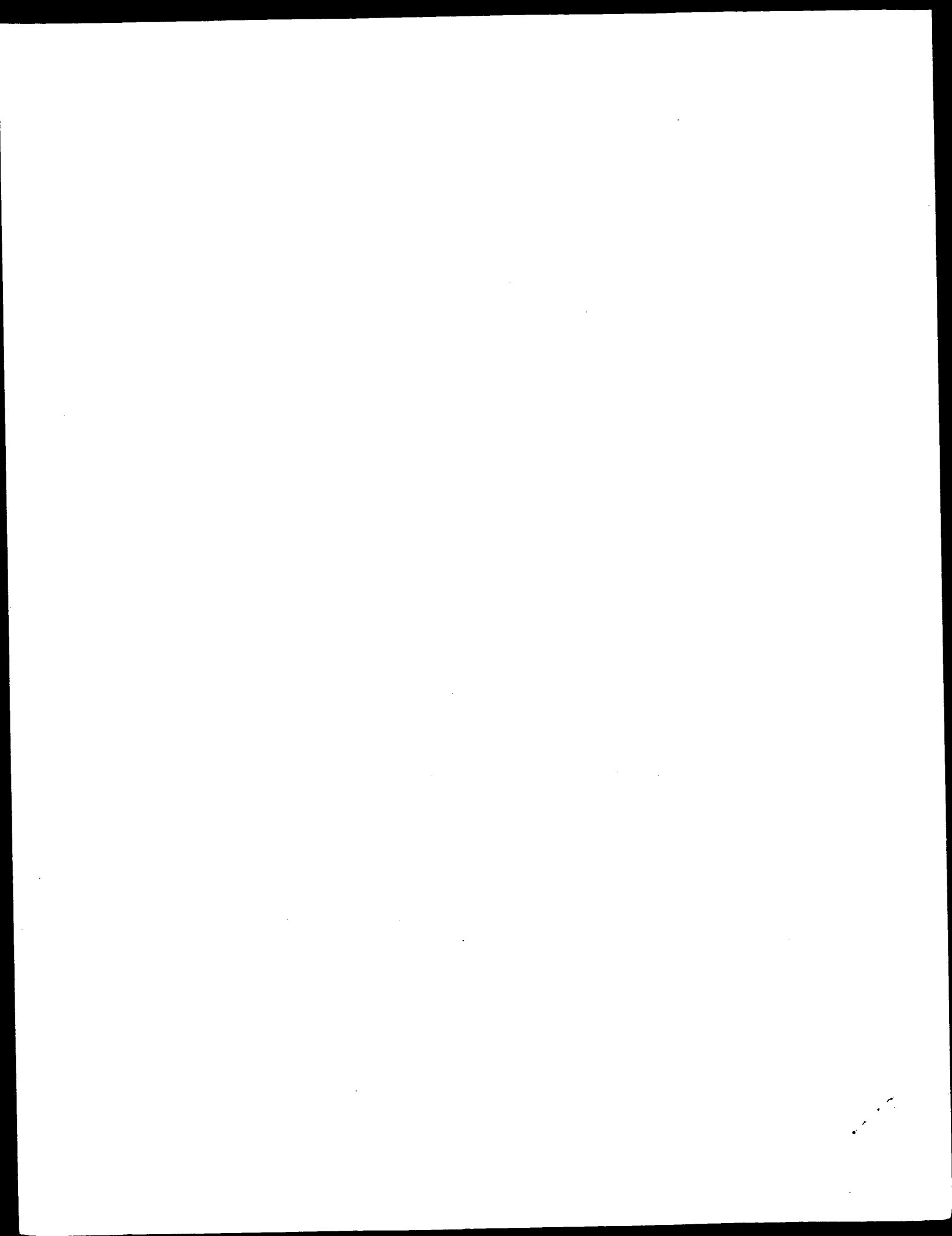
Thu Jun 6 10:09:13 2002

us-08-881-509-5.rst

Page 9

Qy \*1 TGCCTCGTCCTTCTGTTCTGCAAGGCAACTGACCTT 38  
   ||| || ||||| ||||| ||| || |||  
Db 218 TGCTTCCTCCTTCCTGTTCTGTGTGCCAGTGGTCTT 181

Search completed: June 6, 2002, 06:54:29  
Job time: 2692 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 07:00:20 ; Search time 333.99 Seconds  
(without alignments)  
200.484 Million cell updates/sec

Title: US-08-881-509-5

Perfect score:

Sequence: 1 TGCCTCGTCCTTTCTGGTTCTGCAAGGCAACTGACCTTT 39

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters:	3472872
--	---------

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match on

Processing: Minimum Match 0%  
Maximum Match 100%

Maximum Match 100%  
Listing first 45 summaries

```

Database :
1: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	39	100.0		39	19	AAV18707	CDNA for T-cell re
2	27.6	70.8		36	19	AAV18708	CDNA for T-cell re
3	26.2	67.2		1341	19	AAV18705	CDNA for T-cell re
C 4	22.8	58.5		1200	22	AAS06723	Polynucleotide seq
C 5	22.8	58.5		2452	22	AAR44667	Novel protein kina
C 6	22.8	58.5		3735	22	AAI66927	Human protein kina
C 7	22.8	58.5		5499	22	AAI66926	Human protein kina
8	21.6	55.4		1052	22	AAH94558	Human foetal CDNA,
C 9	21.4	54.9		1664	23	AAS89146	DNA encoding novel

C	10	21	53.8	10732	21	AAAL0594	Gene encoding a su
C	11	20.6	52.8	11992	22	AAI199007	Human excretory re
C	12	20.6	52.8	11992	22	AAK80050	Human immune/haema
C	13	20.6	52.8	11992	22	AAI163557	Human kidney relat
C	14	20.4	52.3	1001	21	AAH513555	Human FMO related
C	15	20.4	52.3	1378	23	AAAS81354	DNA encoding novel
C	16	20.4	52.3	4462	21	AAAL1283	Human flavin-conta
C	17	20.4	52.3	26016	19	AAV57272	A. thaliana PUP1 D
C	18	20.2	51.8	1049	21	AAAG97920	Drosophila melanog
C	19	20.2	51.8	4494	23	ABL05292	Renal carcinoma as
C	20	20	51.3	728	16	AAO83561	Human oestrogen re
C	21	20	51.3	465237	24	ABA90193	Cellulose synthase
C	22	19.8	50.8	1000	19	AAV08382	Human cDNA sequenc
C	23	19.8	50.8	2067	22	AAH17966	Human cardiovascu
C	24	19.8	50.8	2533	22	AAAS36156	DNA encoding novel
C	25	19.8	50.8	2967	23	AAAS81075	Human cardiovascu
C	26	19.8	50.8	13176	22	AAAS36157	DNA encoding novel
C	27	19.6	50.3	401	22	AAK95550	Human cardiovascu
C	28	19.6	50.3	401	22	AAK97043	Human neuroregulin g
C	29	19.6	50.3	4482	23	AAST72747	Human neuroregulin g
C	30	19.6	50.3	4482	23	AAAS88068	DNA encoding novel
C	31	19.6	50.3	4974	23	AAST72749	DNA encoding novel
C	32	19.6	50.3	4974	23	AAAS88070	DNA encoding novel
C	33	19.4	49.7	435	21	AAAS51458	DNA encoding novel
C	34	19.4	49.7	1517	23	AAAS91336	PCpB-DFTAEAGGVR-hp
C	35	19.4	49.7	1586	21	AAAF18247	DNA encoding novel
C	36	19.4	49.7	1614	23	AAAS1161	Lung cancer associ
C	37	19.4	49.7	3174	23	ABLI17583	DNA encoding novel
C	38	19.4	49.7	5616	23	ABLI17582	Drosophila melanog
C	39	19.2	49.2	393	21	AAAC06405	Human secreted pro
C	40	19.2	49.2	397	22	AAAL20375	Human breast cance
C	41	19.2	49.2	750	21	AACT77592	Human ORFX ORF3147
C	42	19.2	49.2	1636	22	AAAL60336	Human polynucleoti
C	43	19.2	49.2	1650	22	AAI58850	Human polynucleoti
C	44	19.2	49.2	1788	20	AAAS60777	Soybean cytochrome
C	45	19.2	49.2	3172	22	AAAS34789	DNA encoding novel

## ALIGNMENTS

RESULT	1
NAV18707	
ID	AAV18707 standard; cDNA; 39 BP.
XX	
AC	AAV18707;
XX	
DT	26-JUN-1998 (first entry)
XX	
DE	cDNA for T-cell receptor CDR3 alpha-region.

Gene encoding a su  
Human excretory re  
Human Immune/haema  
Human kidney relat  
Human PWO related  
DNA encoding novel  
Rat MAGUIN 2 CDNA  
Human flavin-CDNA  
A. thaliana pup1 D  
Drosophila melanog  
Renal carcinoma as  
Human oestrogen re  
Cellulose synthase  
Human cDNA sequenc  
Human cardiovascular  
DNA encoding novel  
Human cardiovascular  
Human neuregulin g  
Human neuregulin g  
DNA encoding novel  
DNA encoding novel  
DNA encoding novel  
PCpB-DFLAEGGGVR-hp  
DNA encoding novel  
Lung cancer associ  
DNA encoding novel  
DNA encoding novel  
Drosophila melanog  
Drosophila melanog  
Human secreted pro  
Human breast cancer  
Human ORF3147  
Human polynucleoti  
Human polynucleoti  
Soybean cytochrome  
DNA encoding novel

DR WPI: 1998-053442/06.  
 DR P-PSDB: AAW47590.

XX Human T-cell receptor nucleic acids and poly:peptide(s) - for  
 PT diagnosis or therapy, especially of renal cell carcinoma  
 XX  
 PS Example 1; Page 17; 30pp; German.

XX The present sequence encodes the CDR3 alpha-region of a human  
 CC T-cell receptor (TCR), which can be used in the diagnosis,  
 CC monitoring, prevention and therapy of a tumour disease,  
 CC specifically renal cell carcinoma.

XX Sequence 39 BP; 5 A; 12 C; 8 G; 14 T; 0 other;

Query Match 100.0%; Score 39; DB 19; Length 39;

Best Local Similarity 100.0%; Pred. No. 5.5e-07;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCCTCGTCTTCGTTCTGCAAGGCAACTGACCTTT 39  
 ||| |  
 Db 1 tgcctcgctcttctggttctgcaaggcaactgacctt 39

# RESULT 2

AAV18708  
 ID AAV18708 standard; cDNA; 36 BP.

XX AAV18708;

DT 26-JUN-1998 (first entry)

XX cDNA for T-cell receptor CDR3 alpha-region.

XX Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;  
 KW prevention; therapy; tumour disease; renal cell carcinoma;  
 KW CDR3; ds.

OS Homo sapiens.

XX Key Location/Qualifiers  
 FH mat\_peptide 1..36  
 FT /\*tag= a

XX DE19625191-Al.

XX 02-JAN-1998.

XX 24-JUN-1996; 96DE-1025191.

XX 24-JUN-1996; 96DE-1025191.

XX (BOEF ) BOEHRINGER MANNHEIM GMBH.

XX Schendel D;

XX WPI: 1998-053442/06.  
 DR P-PSDB: AAW47591.

XX Human T-cell receptor nucleic acids and poly:peptide(s) - for  
 PT diagnosis or therapy, especially of renal cell carcinoma

XX Example 1; Page 17; 30pp; German.

XX The present sequence encodes the CDR3 alpha-region of a human  
 CC T-cell receptor (TCR), which can be used in the diagnosis,  
 CC monitoring, prevention and therapy of a tumour disease,  
 CC specifically renal cell carcinoma.

XX Sequence 36 BP; 6 A; 11 C; 8 G; 11 T; 0 other;

Query Match 70.8%; Score 27.6; DB 19; Length 36;  
 Best Local Similarity 88.2%; Pred. No. 0.032;  
 Matches 30; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 CGTCTCTTCGTTCTGCAAGGCAACTGACCTTT 39  
 ||| |  
 Db 3 cctcgctactggttctgcaaggcaactgaccttt 36

# RESULT 3

AAV18705  
 ID AAV18705 standard; cDNA; 1341 BP.

XX AAV18705;

DT 26-JUN-1998 (first entry)

XX cDNA for T-cell receptor alpha-chain.

XX Alpha-chain; human; T-cell receptor; TCR; diagnosis; monitoring;  
 KW prevention; therapy; tumour disease; renal cell carcinoma; ds.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FH CDS 1..804  
 FT /\*tag= a

FT sig\_peptide 1..54

FT /\*tag= b

FT mat\_peptide 55..801

FT /\*tag= c

FT /product= T-cell\_receptor\_alpha-chain

XX DE19625191-Al.

XX 02-JAN-1998.

XX 24-JUN-1996; 96DE-1025191.

XX 24-JUN-1996; 96DE-1025191.

XX (BOEF ) BOEHRINGER MANNHEIM GMBH.

XX Schendel D;

XX WPI: 1998-053442/06.

XX P-PSDB: AAW47588.

XX Human T-cell receptor nucleic acids and poly:peptide(s) - for  
 PT diagnosis or therapy, especially of renal cell carcinoma

XX Example 1; Pages 11-13; 30pp; German.

XX The present sequence encodes the alpha-chain of a human T-cell  
 CC receptor (TCR), which can be used in the diagnosis, monitoring,  
 CC prevention and therapy of a tumour disease, specifically renal  
 CC cell carcinoma.

XX Sequence 1341 BP; 331 A; 365 C; 294 G; 351 T; 0 other;

# Query Match

Best Local Similarity 67.2%; Score 26.2; DB 19; Length 1341;

Matches 31; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 TGCCTCTCTCTTCTGTTCTGCAAGGCAACTGACCTTT 39  
 ||| |  
 Db 310 tactcgctcgtggtggtctgtctgcaaggcaactgaccttt 348

# RESULT 4

AAS06723/C  
 ID AAS06723 standard; cDNA; 1200 BP.



XX AAS06723;  
 AC  
 DT 12-SEP-2001 (first entry)  
 XX  
 DE Polynucleotide sequence encoding human protein kinase #23.  
 XX  
 KW Human; protein kinase; PKK; STK; cancer; cardiovascular disease;  
 KW metabolic disorder; immune related disease; neurological disorder;  
 KW neurodegenerative disorder; inflammatory disorder; infectious disease;  
 KW reproductive disorder; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200138503-A2.  
 XX  
 PD 31-MAY-2001.  
 XX  
 PF 22-NOV-2000; 2000WO-US32085.  
 XX  
 PR 24-NOV-1999; 99US-0167482.  
 XX  
 PA (SUGF-) SUGEN INC.  
 XX  
 PI Plowman GD, Whyte D, Manning G, Sudarsanam S, Martinez R;  
 PI Flanagan P, Clary D;  
 XX  
 DR WPI; 2001-343950/36.  
 DR P-PSDB; AAU03523.  
 XX  
 PT Nucleic acids encoding human kinase polypeptides, useful for preventing  
 PT diagnosing and/or treating e.g. cancer, immune, cardiovascular and  
 PT neuronal-associated diseases, and microbial infections -  
 XX  
 PS Example 1; Figure 1; 433pp; English.  
 XX  
 CC AAS06701-AAS06757 encode for novel human protein kinases #1-57. The  
 CC novel protein kinases have been identified as members of the tyrosine  
 CC or serine/threonine kinase (PTK and STK) families. The polynucleotides  
 CC encoding protein kinases and the polypeptides may be used in the  
 CC prevention, diagnosis and treatment of diseases associated with  
 CC inappropriate kinase expression. For example, they may be used to treat  
 CC cancers (especially cancers of haematopoietic origin), cardiovascular  
 CC disease (e.g. atherosclerosis), metabolic disorders (e.g. diabetes),  
 CC immune related diseases (e.g. rheumatoid arthritis), neurological  
 CC disorders (e.g. schizophrenia), inflammatory disorders (e.g. asthma), infectious  
 CC disease (e.g. HIV) and reproductive disorders (e.g. infertility).  
 CC Additionally, polynucleotides encoding protein kinases may be  
 CC used for gene therapy and as DNA probes in diagnostic assays.  
 CC The protein kinase polypeptides may be used as antigens in the production  
 CC of antibodies against the protein kinases and in assays to identify  
 CC modulators of protein kinase expression and activity.  
 XX  
 SQ Sequence 1200 BP; 333 A; 245 C; 299 G; 320 T; 3 other;

Query Match 58.5%; Score 22.8; DB 22; Length 1200;  
 Best Local Similarity 79.4%; Pred. No. 7;  
 Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 3 CCTCGTCCTTCTGTTCTGCAAGGCAACTGACC 36  
 III IIIIIII I IIII I IIIII IIIII  
 Db 697 CCTGTGCTTTATTTTCTCCAGGCGAGCTGACC 664

RESULT 5  
 AAF44667/c  
 ID AAF44667 standard; CDNA; 2452 BP.  
 XX  
 AC AAF44667;  
 XX  
 DT 27-MAR-2001 (first entry)

Novel protein kinase cDNA, SEQ ID NO: 47.  
 Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;  
 immunosuppressive; cardiant; renal; antiinflammatory; antiasthmatic;  
 dermatological; antidiabetic; antifertility; gene therapy; vaccine;  
 immune disorder; cardiovascular disease; neurodegenerative disease;  
 cancer; autoimmune disorder; stroke; inflammatory bowel disease;  
 inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.  
 OS Homo sapiens.  
 XX  
 PN WO200073469-A2.  
 XX  
 PD 07-DEC-2000.  
 XX  
 PF 26-MAY-2000; 2000WO-US14842.  
 XX  
 PR 28-MAY-1999; 99US-0136503.  
 XX  
 PA (SUGF-) SUGEN INC.  
 XX  
 PI Plowman GD, Martinez R, Whyte D, Sudarsanam S;  
 PI WPI; 2001-032161/04.  
 DR P-PSDB; AAB65640.  
 XX  
 PT Nucleic acids encoding kinase polypeptides, useful for diagnosing and  
 PT treating immune-related diseases and disorders, cardiovascular disease,  
 PT neurodegenerative diseases and/or cancers -  
 XX  
 PS Example 4; Fig 2; 310pp; English.  
 XX  
 CC The present sequence encodes a novel protein kinase. The nucleic acids  
 CC and the protein kinases they encode may be used in the treatment and  
 CC diagnosis of diseases associated with inappropriate kinase expression  
 CC such as immune-related diseases and disorders, cardiovascular disease,  
 CC neurodegenerative diseases and/or cancers. The nucleic acids and  
 CC complementary sequences may also be used as DNA probes in diagnostic  
 CC assays. The kinase polypeptides may be used as antigens in the production  
 CC of antibodies of kinase expression and activity. Anti-kinase antibodies  
 CC and kinase antagonists may also be used to down regulate kinase  
 CC expression and activity. Diseases related to kinase expression and  
 CC activity include rheumatoid arthritis, atherosclerosis, autoimmune  
 CC disorders, complications of organ transplantation, myocardial infarction,  
 CC oxidative-stress related disorders, strokes, renal failure,  
 CC chronic inflammatory pelvic disease, chronic inflammatory bowel disease,  
 CC osteoarthritis, psoriasis, rhinitis, multiple sclerosis, asthma,  
 CC reproductive disorders.  
 XX  
 SQ Sequence 2452 BP; 676 A; 529 C; 624 G; 622 T; 1 other;

Query Match 58.5%; Score 22.8; DB 22; Length 2452;  
 Best Local Similarity 79.4%; Pred. No. 8.1;  
 Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 3 CCTCGTCCTTCTGTTCTGCAAGGCAACTGACC 36  
 III IIIIIII I IIII I IIIII IIIII  
 Db 1102 CCTGTGCTTTATTTTCTCCAGGCGAGCTGACC 1069

RESULT 6  
 AAI66827/c  
 ID AAI66827 standard; CDNA; 3735 BP.  
 XX  
 AC AAI66827;  
 XX  
 DT 07-JAN-2002 (first entry)  
 XX  
 DE Human protein kinase polypeptide 23546 coding sequence.  
 XX

Thu Jun 6 10:09:10 2002

Protein kinase: 3714; 16742; 23546; 13887; cancer; bone disorder; human; cytostatic; antiinflammatory; immunosuppressive; cardiant; hepatotrophic; gene therapy; ss.

Homo sapiens.

Key Location/Qualifiers  
CDS 1..3755  
/tag= a  
/product= "protein kinase 23546"

WO200173050-A2.

04-OCT-2001.

23-MAR-2001; 2001WO-US09483.

24-MAR-2000; 2000US-191846P.

(MILL-) MILLENNIUM PHARM INC.

Meyers R;

WPI; 2001-611632/70.

P-PSDB; AAG65766.

New human protein kinase polypeptides, 3714, 16742, 23546 and 13887, useful in diagnosis of cancer or cellular proliferation or differentiation disorders and to screen for polypeptide modulators useful to treat such conditions

Claim 1; Fig 7A-G; 169pp; English.

The invention provides novel human protein kinase polypeptides, 3714, 16742, 23546 and 13887 and nucleic acid molecules encoding them. The protein kinase polypeptides can be expressed by standard recombinant methodology. 3714, 16742, 23546 or 13887 nucleic acids and polypeptides are useful for diagnostic and screening methods to identify subjects (at risk of) having cancer or cellular proliferation and/or differentiation disorders. 3714, 16742, 23546 or 13887 nucleic acids, polypeptides and modulators are useful for the treatment of cancer, particularly colon cancer or cellular proliferation and/or differentiation disorders. Other disorders associated with 3714, 16742, 23546 or 13887 expression or activity that can be treated include bone related disorders, inflammatory disorders, autoimmune diseases, cardiovascular disorders and liver diseases. The present sequence represents a human protein kinase polypeptide 23546 coding sequence.

Sequence 3735 BP; 1088 A; 878 C; 883 G; 886 T; 0 other;

Query Match 58.5%; Score 22.8; DB 22; Length 3735;  
Best Local Similarity 79.4%; Pred. No. 8.9;  
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 CCTGCTCCTTCTGGTTCTCGAAGGCAACTGACC 36  
||| ||||| | |||| | |||| | |||| | |||| |  
Db 697 CCTGCTCCTTATTTTCTCCAGGCGAGCTGACC 664

RESULT 7

AAI66826/c

ID AAI66826 standard; cDNA; 5499 BP.

XX AC AAI66826;

XX DT 07-JAN-2002 (first entry)

XX DE Human protein kinase polypeptide 23546 encoding cDNA.

XX DE Protein kinase; 3714; 16742; 23546; 13887; cancer; bone disorder; human;  
KW cytostatic; antiinflammatory; immunosuppressive; cardiant; hepatotrophic;  
KW gene therapy; ss.

XX OS Homo sapiens.

Key Location/Qualifiers  
5'UTR 1..316  
/tag= a

FT 317..4051  
/tag= b

FT /product= "protein kinase 23546"

FT /note= "coding sequence specifically claimed"

FT 4052..5499

FT /tag= c

PN WO200173050-A2.

XX 04-OCT-2001.

XX 23-MAR-2001; 2001WO-US09483.

XX 24-MAR-2000; 2000US-191846P.

XX (MILL-) MILLENNIUM PHARM INC.

XX Meyers R;

PI WPI; 2001-611632/70.

XX P-PSDB; AAG65766.

XX New human protein kinase polypeptides, 3714, 16742, 23546 and 13887, useful in diagnosis of cancer or cellular proliferation or differentiation disorders and to screen for polypeptide modulators useful to treat such conditions

Claim 1; Fig 7A-G; 169pp; English.

The invention provides novel human protein kinase polypeptides, 3714, 16742, 23546 and 13887 and nucleic acid molecules encoding them. The protein kinase polypeptides can be expressed by standard recombinant methodology. 3714, 16742, 23546 or 13887 nucleic acids and polypeptides are useful for diagnostic and screening methods to identify subjects (at risk of) having cancer or cellular proliferation and/or differentiation disorders. 3714, 16742, 23546 or 13887 nucleic acids, polypeptides and modulators are useful for the treatment of cancer, particularly colon cancer or cellular proliferation and/or differentiation disorders. Other disorders associated with 3714, 16742, 23546 or 13887 expression or activity that can be treated include bone related disorders, inflammatory disorders, autoimmune diseases, cardiovascular disorders and liver diseases. The present sequence represents a human protein kinase polypeptide 23546 encoding cDNA.

Sequence 5499 BP; 1600 A; 1215 C; 1258 G; 1426 T; 0 other;

Query Match 58.5%; Score 22.8; DB 22; Length 5499;  
Best Local Similarity 79.4%; Pred. No. 9.7;  
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 CCTGCTCCTTCTGGTTCTCGAAGGCAACTGACC 36  
||| ||||| | |||| | |||| | |||| | |||| |  
Db 1013 CCTGCTCCTTATTTTCTCCAGGCGAGCTGACC 980

RESULT 8

AAH94558

ID AAH94558 standard; cDNA; 1052 BP.

XX AC AAH94558;

XX DT 05-OCT-2001 (first entry)

XX DE Human foetal cDNA, SEQ ID NO: 1245.

XX KW Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;

KW neotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;  
 KW gene therapy; antisense therapy; cancer; immune disorder;  
 KW growth disorder; osteoporosis; thrombolytic disorder;  
 KW nervous system disorder; inflammation; expressed sequence tag; EST; ss.  
 OS Homo sapiens.  
 XX WO200155339-A2.  
 PN 02-AUG-2001.  
 PD 25-JAN-2001; 2001WO-US02723.  
 XX 25-JAN-2000; 2000US-0491404.  
 PR 15-SEP-2000; 2000US-0663870.  
 PR 06-NOV-2000; 2000US-0707351.  
 XX (HYSE-) HYSEQ INC.  
 PA Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;  
 PI Liu C, Asundi V, Zhou P, Werhman T;  
 XX WPI; 2001-465571/50.  
 DR P-PSDB; AAM06883.

XX Novel fetal proteins useful for the treatment and diagnosis of diseases  
 PT associated with dysfunction of the protein e.g. cancers, immune  
 PT disorders, growth disorders, thrombolytic disorders, nervous system  
 PT disorders and inflammation -  
 XX Example 3; Page 705; 715pp; English.  
 PS The invention relates to novel foetal polypeptides encoded by  
 CC polynucleotides comprising one of 477 sequences fully defined in the  
 CC specification. The foetal polynucleotides and polypeptides are  
 CC useful in the treatment and diagnosis of diseases such as cancers,  
 CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic  
 CC disorders, nervous system disorders and inflammation. The present  
 CC sequence is a full length cDNA which was assembled using expressed  
 CC sequence tags (ESTs) found to be expressed in human foetal tissue  
 CC cDNA libraries as seeds.  
 XX Sequence 1052 BP; 299 A; 188 C; 248 G; 316 T; 1 other;

Query Match 55.4%; Score 21.6; DB 22; Length 1052;  
 Best Local Similarity 75.0%; Pred. No. 22;  
 Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 CCTCGCTCTTCTGTTCTGCAAGGCAACTGACCTT 38  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 407 cttaactctttctgttcttctgtagcattgacatt 442

RESULT 9  
 AAS89146/c  
 ID AAS89146 standard; cDNA; 1664 BP.  
 XX AAS89146;  
 AC AAS89146;  
 XX 13-FEB-2002 (first entry)  
 DT DNA encoding novel human diagnostic protein #24950.  
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 KW Homo sapiens.  
 OS WO200175067-A2.  
 PN 11-OCT-2001.  
 PD 11-OCT-2001.  
 XX

PF 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 PI P-PSDB; ABG24959.  
 DR New isolated polynucleotide and encoded polypeptides, useful in  
 DR diagnostics, forensics, gene mapping, identification of mutations  
 DR responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PT Claim 1; SEQ ID No 24950; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 1664 BP; 407 A; 499 C; 425 G; 333 T; 0 other;

Query Match 54.9%; Score 21.4; DB 23; Length 1664;  
 Best Local Similarity 71.8%; Pred. No. 29;  
 Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 TGCCTCGTCTTCTGTTCTGCAAGGCAACTGACCTTT 39  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 1048 TTCTCGTCATTCTGCGCCAGCTGCTCCTGCTGCTTT 1010

RESULT 10  
 AAA10594/c  
 ID AAA10594 standard; DNA; 10732 BP.  
 XX AAA10594;  
 AC AAA10594;  
 XX 29-JUN-2000 (first entry)  
 DT Gene encoding a subunit of cellulose synthase.  
 DE Cellulose synthase; cellulose production; increase yield; ds.  
 KW Vigna angularis.  
 XX JP2000060568-A.  
 PN 29-FEB-2000.  
 PD 29-FEB-2000.  
 XX 26-AUG-1998; 98JP-0239998.  
 XX



PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249246.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249267.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	05-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	11-DEC-2000;	2000US-0251990.
PR	03-JAN-2001;	2000US-0254037.
PR	03-JAN-2001;	2001US-0259678.

CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at [ftp:wipo.int/pub/published\\_pct\\_sequences](ftp:wipo.int/pub/published_pct_sequences).  
XX

Query Match	52.8%;	Score 20.6;	DB 22;	Length 11992;
Best Local Similarity	74.3%;	Pred. No. 95;		

OY 3 CCTCGTCTTTCTGGTTCTGCAAGGCAACTGACCT 37  
||| ||| ||| ||| ||| ||| ||| |||  
db 6848 CCTTCACCTCTATGGTCCCTCCAATGCAACTGA 6814

RESULT 12  
AAK80050/c  
IID AAK80050 standard; DNA; 11992 BP.

PR	22-AUG-2000;	2000US-0226681.	PR	17-NOV-2000;	2000US-0249207.
PR	22-AUG-2000;	2000US-0226688.	PR	17-NOV-2000;	2000US-0249208.
PR	22-AUG-2000;	2000US-0227182.	PR	17-NOV-2000;	2000US-0249209.
PR	22-AUG-2000;	2000US-0227009.	PR	17-NOV-2000;	2000US-0249210.
PR	30-AUG-2000;	2000US-0228924.	PR	17-NOV-2000;	2000US-0249211.
PR	01-SEP-2000;	2000US-0229287.	PR	17-NOV-2000;	2000US-0249212.
PR	01-SEP-2000;	2000US-0229343.	PR	17-NOV-2000;	2000US-0249213.
PR	01-SEP-2000;	2000US-0229344.	PR	17-NOV-2000;	2000US-0249214.
PR	01-SEP-2000;	2000US-0229345.	PR	17-NOV-2000;	2000US-0249215.
PR	05-SEP-2000;	2000US-0229509.	PR	17-NOV-2000;	2000US-0249216.
PR	05-SEP-2000;	2000US-0229513.	PR	17-NOV-2000;	2000US-0249217.
PR	06-SEP-2000;	2000US-0230437.	PR	17-NOV-2000;	2000US-0249218.
PR	06-SEP-2000;	2000US-0230438.	PR	17-NOV-2000;	2000US-0249244.
PR	08-SEP-2000;	2000US-0231243.	PR	17-NOV-2000;	2000US-0249245.
PR	08-SEP-2000;	2000US-0231244.	PR	17-NOV-2000;	2000US-0249264.
PR	08-SEP-2000;	2000US-0231245.	PR	17-NOV-2000;	2000US-0249265.
PR	08-SEP-2000;	2000US-0231413.	PR	17-NOV-2000;	2000US-0249297.
PR	08-SEP-2000;	2000US-0231414.	PR	17-NOV-2000;	2000US-0249299.
PR	08-SEP-2000;	2000US-0232080.	PR	17-NOV-2000;	2000US-0249300.
PR	08-SEP-2000;	2000US-0232081.	PR	01-DEC-2000;	2000US-0250160.
PR	12-SEP-2000;	2000US-0231968.	PR	01-DEC-2000;	2000US-0250391.
PR	14-SEP-2000;	2000US-0232397.	PR	05-DEC-2000;	2000US-0251030.
PR	14-SEP-2000;	2000US-0232398.	PR	05-DEC-2000;	2000US-0251988.
PR	14-SEP-2000;	2000US-0232399.	PR	05-DEC-2000;	2000US-0256719.
PR	14-SEP-2000;	2000US-0232400.	PR	06-DEC-2000;	2000US-0251479.
PR	14-SEP-2000;	2000US-0232401.	PR	08-DEC-2000;	2000US-0251856.
PR	14-SEP-2000;	2000US-0233063.	PR	08-DEC-2000;	2000US-0251868.
PR	14-SEP-2000;	2000US-0233064.	PR	08-DEC-2000;	2000US-0251869.
PR	14-SEP-2000;	2000US-0233065.	PR	08-DEC-2000;	2000US-0251989.
PR	21-SEP-2000;	2000US-0234223.	PR	08-DEC-2000;	2000US-0251990.
PR	21-SEP-2000;	2000US-0234274.	PR	11-DEC-2000;	2000US-0254097.
PR	25-SEP-2000;	2000US-0234997.	PR	05-JAN-2001;	2001US-0259678.
PR	25-SEP-2000;	2000US-0234998.	XX		
PR	26-SEP-2000;	2000US-0235484.	PA	(HUMA-) HUMAN GENOME SCI INC.	
PR	27-SEP-2000;	2000US-0235834.	XX		
PR	27-SEP-2000;	2000US-0235836.	PI	Rosen CA, Barash SC, Ruben SM;	
PR	29-SEP-2000;	2000US-0236327.	XX	WPI; 2001-483426/52.	
PR	29-SEP-2000;	2000US-0236367.	DR		
PR	29-SEP-2000;	2000US-0236368.	XX		
PR	29-SEP-2000;	2000US-0236369.	PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
PR	29-SEP-2000;	2000US-0236370.	PT	useful for preventing, diagnosing and/or treating cancers and	
PR	02-OCT-2000;	2000US-0236602.	XX	metastasis -	
PR	02-OCT-2000;	2000US-0237037.	PS	Disclosure; SEQ ID NO 34862; 3071pp + Sequence Listing; English.	
PR	02-OCT-2000;	2000US-0237038.	XX		
PR	02-OCT-2000;	2000US-0237039.	CC	AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)	
PR	02-OCT-2000;	2000US-0237040.	CC	amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic	
PR	13-OCT-2000;	2000US-0239835.	CC	activity, and can be used in gene therapy and vaccine production. (I)	
PR	20-OCT-2000;	2000US-0240960.	CC	proteins and polynucleotides may be used in the prevention, diagnosis and	
PR	20-OCT-2000;	2000US-02411785.	CC	treatment of diseases associated with inappropriate (I) expression. For	
PR	20-OCT-2000;	2000US-0241786.	CC	example, they may be used to treat disorders associated with decreased	
PR	20-OCT-2000;	2000US-0241787.	CC	expression by rectifying mutations or deletions in a patient's genome	
PR	20-OCT-2000;	2000US-0241808.	CC	that affect the activity of (I) by expressing inactive proteins or to	
PR	20-OCT-2000;	2000US-0241809.	CC	supplement the patients own production of (I). Additionally, (I)	
PR	20-OCT-2000;	2000US-0241826.	CC	polynucleotides may be used to produce the secreted (I), by inserting	

RESULT 13  
AAI63357/C  
ID AAI63357 standard; DNA; 11992 BP.  
XX AC AAI63357;  
XX DT 22-OCT-2001 (first entry)  
XX DE Human kidney related polynucleotide SEQ ID NO 672.  
XX KW Human; kidney antigen; immunosuppressive; antiarthritic; antirheumatic;  
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
KW neurotropic; neuroprotective; antibacterial; virucide; fungicide;  
KW ophthalmological; antiallergic; hepatotropic; antidiabetic;  
KW antinflammatory; antitumor; vulnery; anticonvulsant; antiparasitic;  
KW gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; ds.  
XX OS Homo sapiens.  
XX PN WO200155323-A2.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01343.  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216547.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225271.  
PR 14-AUG-2000; 2000US-0225275.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 18-AUG-2000; 2000US-0225759.  
PR 22-AUG-2000; 2000US-0226278.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0227009.  
PR 01-SEP-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 08-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.

17-NOV-2000; 2000US-0249264.  
17-NOV-2000; 2000US-0249265.  
17-NOV-2000; 2000US-0249297.  
17-NOV-2000; 2000US-0249299.  
17-NOV-2000; 2000US-0249300.  
01-DEC-2000; 2000US-0250160.  
01-DEC-2000; 2000US-0250391.  
05-DEC-2000; 2000US-0251030.  
05-DEC-2000; 2000US-0251988.  
05-DEC-2000; 2000US-0256719.  
06-DEC-2000; 2000US-0251479.  
08-DEC-2000; 2000US-0251856.  
08-DEC-2000; 2000US-0251868.  
08-DEC-2000; 2000US-0251869.  
08-DEC-2000; 2000US-0251989.  
08-DEC-2000; 2000US-0251990.  
11-DEC-2000; 2000US-0254097.  
05-JAN-2001; 2001US-0259678.  
XX XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-488784/53.  
DR New isolated nucleic acids and polypeptides, useful for diagnosing,  
PT treating and/or preventing human diseases and disorders -  
XX Disclosure; SEQ ID NO 672; 564pp + Sequence Listing; English.  
XX  
XX The invention relates to novel kidney related polynucleotides  
CC (AAI62971-AAI63793) and the encoded polypeptides (AAI62971-AAI63793) and  
CC collectively known as kidney antigens and the use of such kidney antigens  
CC for detecting disorders of the kidney, especially kidney cancer and  
CC kidney cancer metastases. The polynucleotides and proteins are also  
CC useful for preventing, treating or ameliorating medical conditions  
CC e.g. by protein or gene therapy. The genes are isolated from a range  
CC of human tissues disclosed in the specification. The nucleic acids,  
CC proteins, antibodies and (ant)agonists are useful in the diagnosis,  
CC treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer,  
CC and other cancers of the adrenal gland, bone, bone marrow, breast,  
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders  
CC e.g. Addison's disease, allergies, autoimmune haemolytic anaemia,  
CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple  
CC sclerosis, rheumatoid arthritis and ulcerative colitis;  
CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound  
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;  
CC and (f) infectious diseases such as viral, bacterial, fungal and  
CC parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX

```

Query Match      52.8%; Score 20.6; DB 22; Length 11992;
Best Local Similarity 74.3%; Pred. No. 95;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY      3 CCTCGTCCTTTCTGTGTTCTCAAGGCAACTGAAC T 37
          ||| ||| ||| ||| ||| ||| ||| |||
Db      6848 CCTTCCACTCTATGGTCTCCAATGCAACTGAACT 6814

RESULT 14
AAH51555
ID AAH51555 standard; DNA; 1001 BP.
XX XX
AC AC
AAH51555;
XX XX
DT DT
29-AUG-2001 (first entry)
XX XX
```



Query Match	52.3%	Score 20.4;	DB 23;	Length 1378;
Best Local Similarity	71.1%	Pred. No. 73;		
Matches 27;	Conservative	0; Mismatches	11;	Indels 0; Gaps 0;
2	GCCTCGCTGCTTTCTGGTTCTGCAAGCAACTGACCTTT	39		
286	ggccggcctctctctgctctggctggcattcccttt	323		

Thu Jun 6 10:09:10 2002

us-08-881-509-5.rng

---

Page 12

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 07:38:13 ; Search time 78.65 seconds  
(without alignments)  
121.802 Million cell updates/sec

Title: US-08-881-509-5

Perfect score: 39

Sequence: 1 TGCCTCGCTCTTCTGCTTCTGCAAGCAACTGACCTTT 39

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.2	49.2	1788	3	US-08-948-564-9
2	19	48.7	11459	4	US-09-454-721A-3
3	18.8	48.2	1713	4	US-09-181-336-12
4	18.8	48.2	1721	2	US-08-922-170B-9
5	18.8	48.2	1721	2	US-08-922-170B-11
6	18.8	48.2	1721	4	US-09-071-739B-1
7	18.8	48.2	1721	4	US-09-071-739B-3
8	18.8	48.2	1721	4	US-09-260-038B-1
9	18.8	48.2	1721	4	US-09-260-038B-3
10	18.8	48.2	1723	4	US-09-181-336-14
11	18.6	47.7	2325	1	US-08-019-870-2
12	18.6	47.7	2325	1	US-08-019-870-4
13	18.6	47.7	2325	1	US-08-019-870-7
14	18.6	47.7	2325	1	US-08-019-870-10
15	18.6	47.7	2325	1	US-08-314-309A-20
16	18.6	47.7	2325	1	US-08-633-760-43
17	18.6	47.7	2325	1	US-08-633-760-45
18	18.6	47.7	2325	1	US-08-633-760-47
19	18.6	47.7	2325	1	US-08-633-760-49
20	18.6	47.7	2325	1	US-08-633-760-51
21	18.6	47.7	2373	1	US-08-019-870-9
22	18.6	47.7	2847	1	US-07-747-901A-2
23	18.6	47.7	2847	1	US-07-935-312-2
24	18.4	47.2	2220	6	5171850-1
25	18.4	47.2	2585	4	US-09-008-697A-7
26	18.4	47.2	5496	3	US-08-600-982-23
27	18.4	47.2	5496	5	PCT-US94-10261A-23

28	18.2	46.7	301	4	US-09-439-313-250	Sequence 250, App
29	18.2	46.7	420	4	US-08-943-731-198	Sequence 198, App
30	18.2	46.7	1383	1	US-08-289-709-2	Sequence 2, Appli
31	18.2	46.7	1383	1	US-08-602-656-2	Sequence 2, Appli
32	18.2	46.7	1474	1	US-08-465-980-1	Sequence 1, Appli
33	18.2	46.7	1474	2	PCT-US95-07093-1	Sequence 1, Appli
34	18.2	46.7	1474	5	PCT-US95-07093-1	Sequence 1, Appli
35	18.2	46.7	1828	2	US-08-888-497-29	Sequence 29, Appli
36	18.2	46.7	1828	4	US-09-362-230-29	Sequence 29, Appli
37	18.2	46.7	1828	5	PCT-US94-07926-29	Sequence 29, Appli
38	18.2	46.7	3083	4	US-08-956-657-1	Sequence 1, Appli
39	18.2	46.7	6842	1	US-08-727-034-5	Sequence 5, Appli
40	18.2	46.7	6842	1	US-08-727-034-6	Sequence 6, Appli
41	18.2	46.7	20084	4	US-08-943-731-5	Sequence 5, Appli
42	18	46.2	1392	4	US-09-163-444-1	Sequence 1, Appli
43	18	46.2	1735	4	US-09-163-444-3	Sequence 3, Appli
44	18	46.2	2768	1	US-08-321-478-1	Sequence 1, Appli
45	18	46.2	2768	1	US-08-321-478-3	Sequence 3, Appli

## ALIGNMENTS

RESULT 1  
US-08-948-564-9  
; Sequence 9, Application US/08948564  
; Patent No. 6121512  
; GENERAL INFORMATION:  
; APPLICANT: Siminsky, Balazs  
; APPLICANT: Dewey, Ralph E.  
; APPLICANT: Corbin, Frederick T.  
; TITLE OF INVENTION: No. 6121512el Cytochrome P-450 Constructs and  
; TITLE OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plants  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Virginia C. Bennett  
; STREET: PO Box 37428  
; CITY: Raleigh  
; STATE: No. 6121512th Carolina  
; COUNTRY: USA  
; ZIP: 27627  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/948,564  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bennett, Virginia C.  
; REGISTRATION NUMBER: 37,092  
; REFERENCE/DOCKET NUMBER: 5051-409  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-854-1400  
; TELEFAX: 919-854-1401  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1788 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 6..1601  
US-08-948-564-9

Query Match 49.2%; Score 19.2; DB 3; Length 1788;  
Best Local Similarity 87.5%; Pred. No. 24;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;



```
;
; APPLICANT: Feinstein
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A POLYPEPTIDE
; TITLE OF INVENTION: HAVING HEPARANASE ACTIVITY AND EXPRESSION OF
; TITLE OF INVENTION: SAME IN TRANSFUSED CELLS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein
; STREET: 2940 Birchtree lane
; CITY: Silver Spring
; STATE: Maryland
; COUNTRY: United States of America
; ZIP: 20906
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; OPERATING SYSTEM: Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/922.170B
; FILING DATE: 2 SEP 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/1
; TELEPHONE: 972-3-5625553
; TELEFAX: 972-3-5625554
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1721
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-922-170B-11
;
; Query Match 48.2%; Score 18.8; DB 2; Length 1721;
; Best Local Similarity 68.4%; Pred. No. 35;
; Matches 26; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
;
; QY 1 TGCCCTCGTCCTTCTCGTTCGCAAGCAACTGACCTT 38
; | |||| |||| |||| |||| |||| || ||||
; DB 273 TTCTCATCTCCCTGGGTCTCCAAAGCTTCGTACCTT 310
;
; RESULT 6
; US-071-739B-1
; Sequence 1, Application US/09071739B
; Patent No. 6177545
; GENERAL INFORMATION:
; APPLICANT: Iris Pecker et al.
; TITLE OF INVENTION: HEPARANASE SPECIFIC MOLECULAR PROBES
; TITLE OF INVENTION: AND THEIR USE IN RESEARCH AND MEDICAL
; TITLE OF INVENTION: APPLICATIONS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 20001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; OPERATING SYSTEM: Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,739B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/922.180
; FILING DATE: September 2, 1997
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/5
; OPERATING SYSTEM: MS DOS version 6.2,
```

```
;
; OPERATING SYSTEM: Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,739B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/922.180
; FILING DATE: September 2, 1997
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/5
; TELEPHONE: 972-3-5625553
; TELEFAX: 972-3-5625554
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1721
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-09-071-739B-1
;
; Query Match 48.2%; Score 18.8; DB 4; Length 1721;
; Best Local Similarity 68.4%; Pred. No. 35;
; Matches 26; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
;
; QY 1 TGCCCTCGTCCTTCTCGTTCGCAAGCAACTGACCTT 38
; | |||| |||| |||| |||| |||| || ||||
; DB 273 TTCTCATCTCCCTGGGTCTCCAAAGCTTCGTACCTT 310
;
; RESULT 7
; US-09-071-739B-3
; Sequence 3, Application US/09071739B
; Patent No. 6177545
; GENERAL INFORMATION:
; APPLICANT: Iris Pecker et al.
; TITLE OF INVENTION: HEPARANASE SPECIFIC MOLECULAR PROBES
; TITLE OF INVENTION: AND THEIR USE IN RESEARCH AND MEDICAL
; TITLE OF INVENTION: APPLICATIONS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 20001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; OPERATING SYSTEM: Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,739B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/922.180
; FILING DATE: September 2, 1997
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/5
```

Thu Jun 6 10:09:10 2002

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-5625553
; TELEFAX: 972-3-5625554
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1721
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: linear
;
US-09-071-739B-3

Query Match 48.2%; Score 18.8; DB 4; Length 1721;
Best Local Similarity 68.4%; Pred. No. 35;
Matches 26; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 TGCCTCGCTCTTCTGCTTCTGCAAGGCAACTGACCTT 38
   ||||| ||||| ||||| ||||| ||||| |||||
Db 273 TTCTCATCTCTCTGCTTCTTCCAAAGCTTCGTACCTT 310

RESULT 8
US-09-260-038B-1
; Sequence 1, Application US/09260038B
; Patent No. 6348344
; GENERAL INFORMATION:
; APPLICANT: Maty Ayal-Hershkovitz et al.
; TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR
; EXPRESSING RECOMBINANT HEPARANASE
; AND METHODS OF PURIFYING SAME
;
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 2001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; an ASCII file
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/260,038B
; FILING DATE: 02-Mar-1999
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/071,618
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 09/071,739
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 08/922,180
; FILING DATE: September 2, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/16
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-5625553
; TELEFAX: 972-3-5625554
; TELEX: <Unknown>
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1721
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
;
US-09-260-038B-1
```

```
Query Match 48.2%; Score 18.8; DB 4; Length 1721;
Best Local Similarity 68.4%; Pred. No. 35;
Matches 26; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 TGCCTCGCTCTTCTGCTTCTGCAAGGCAACTGACCTT 38
   ||||| ||||| ||||| ||||| ||||| |||||
Db 273 TTCTCATCTCTCTGCTTCTTCCAAAGCTTCGTACCTT 310

RESULT 9
US-09-260-038B-3
; Sequence 3, Application US/09260038B
; Patent No. 6348344
; GENERAL INFORMATION:
; APPLICANT: Maty Ayal-Hershkovitz et al.
; TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR
; EXPRESSING RECOMBINANT HEPARANASE
; AND METHODS OF PURIFYING SAME
;
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 2001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; an ASCII file
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/260,038B
; FILING DATE: 02-Mar-1999
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/071,618
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 09/071,739
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 08/922,180
; FILING DATE: September 2, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/16
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-5625553
; TELEFAX: 972-3-5625554
; TELEX: <Unknown>
;
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1721
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
;
US-09-260-038B-3

Query Match 48.2%; Score 18.8; DB 4; Length 1721;
Best Local Similarity 68.4%; Pred. No. 35;
Matches 26; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

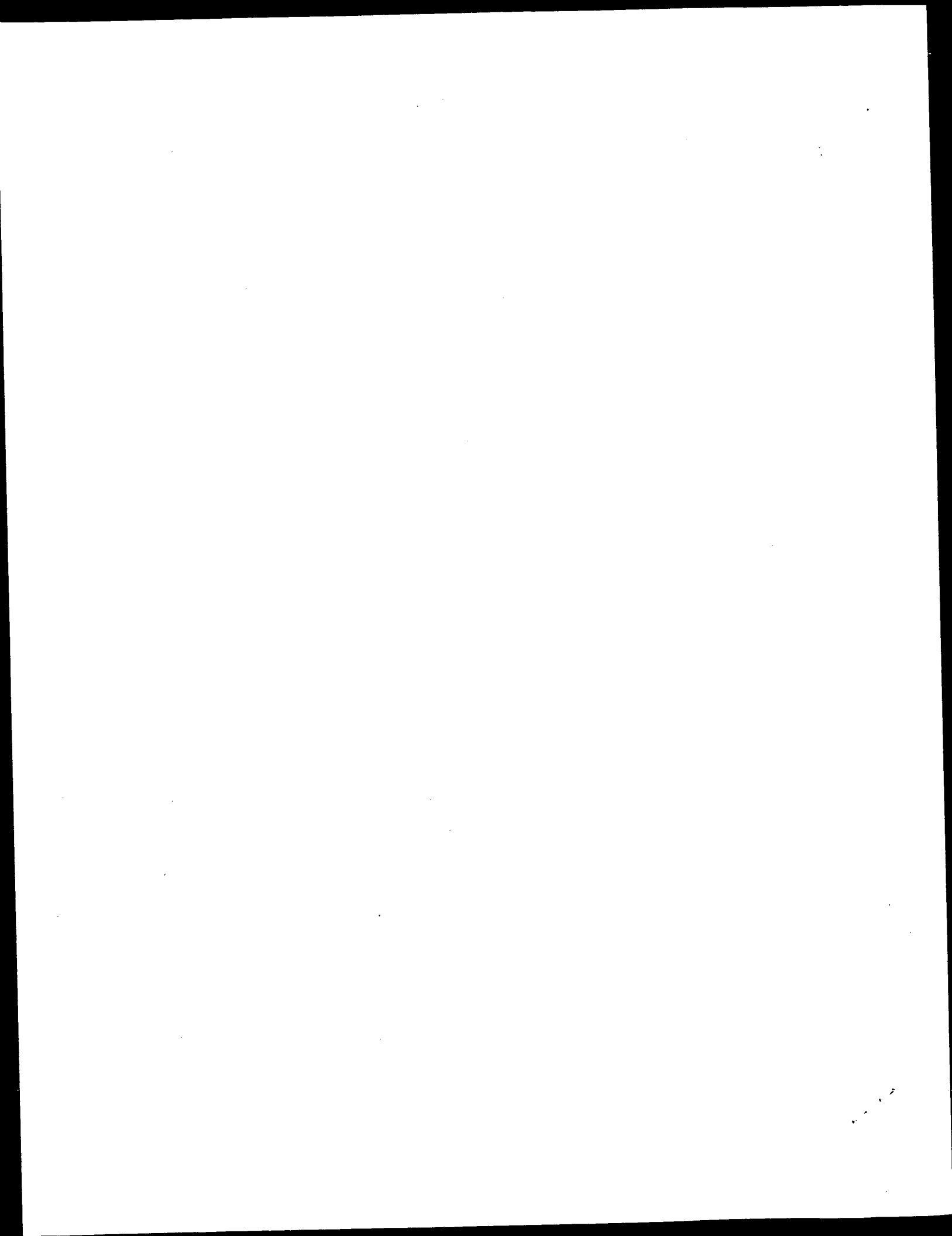
QY 1 TGCCTCGCTCTTCTGCTTCTGCAAGGCAACTGACCTT 38
   ||||| ||||| ||||| ||||| ||||| |||||
Db 273 TTCTCATCTCTCTGCTTCTTCCAAAGCTTCGTACCTT 310
```

RESULT 11  
US-08-019-870-2/c  
Sequence 2, Application US/08019870  
Patent No. 5336613  
GENERAL INFORMATION:  
APPLICANT: NIWA, MINEO  
APPLICANT: YOSHIMASA, SAITO  
APPLICANT: SASAKI, HITOSHI  
APPLICANT: ISHII, YOSHINORI  
TITLE OF INVENTION: A NEW CEPHALOSPORIN C ACYLASE  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/019,870  
FILING DATE: 19930219  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5336613man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 18-791-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:15:16 ; Search time 27.21 Seconds  
(without alignments)  
45.908 Million cell updates/sec

Title: US-08-881-509-6  
Perfect score: 64  
Sequence: 1 CLVLGSGARQLTF 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	60.9	82	2 S35770	T-cell receptor al
2	38	59.4	356	2 AE2729	cytochrome oxidase
3	38	59.4	372	2 F97510	BA483f11.2.1 (cox1
4	38	59.4	803	1 IJCH3	Integrin, band 3.p
5	37	57.8	223	2 T09926	cytidine deaminase
6	37	57.8	337	2 T09925	cytidine deaminase
7	36	56.2	173	1 NKVGGU	core protein p20 -
8	36	56.2	260	2 E84668	hypothetical prote
9	36	56.2	262	2 F87249	enoyl-CoA hydratase
10	36	56.2	315	1 S74441	iron(III) dicitrat
11	36	56.2	347	2 AE0787	cytochrome c-type
12	36	56.2	347	2 AE0787	Cytochrome c-type
13	36	56.2	415	1 W2WLEP	E2 protein - Europ
14	36	56.2	606	2 B87318	hypothetical prote
15	35	54.7	152	2 S36550	E6 protein - human
16	35	54.7	254	2 JC5004	attacin A precursor
17	35	54.7	307	2 T09923	cytidine deaminase
18	35	54.7	336	2 A43260	glyceraldehyde-3-p
19	35	54.7	379	2 JN0013	synaptic vesicle m
20	35	54.7	741	2 AF2297	hypothetical prote
21	35	54.7	905	2 T40015	phosphatidylethano
22	35	54.7	2715	2 T13049	eyelid - fruit fly
23	34.5	53.9	2039	2 S64540	probable calcium c
24	34	53.1	148	2 S36532	E6 protein - human
25	34	53.1	183	2 E85182	hypothetical prote
26	34	53.1	192	2 A64835	probable membrane
27	34	53.1	192	2 F90757	hypothetical prote
28	34	53.1	192	2 D85621	hypothetical prote
29	34	53.1	199	2 S48362	probable membrane

30 34 53.1 208 2 AH2535  
31 34 53.1 256 2 H69747  
32 34 53.1 272 1 G89587  
33 34 53.1 301 2 T52549  
34 34 53.1 305 2 C85878  
35 34 53.1 339 2 E83496  
36 34 53.1 349 2 T06680  
37 34 53.1 388 2 E65054  
38 34 53.1 394 2 T41639  
39 34 53.1 412 2 T41141  
40 34 53.1 419 2 C86223  
41 34 53.1 422 1 A60503  
42 34 53.1 423 2 S56672  
43 34 53.1 423 2 B75503  
44 34 53.1 445 2 B49776  
45 34 53.1 615 2 S77332

## ALIGNMENTS

RESULT 1

S35770  
T-cell receptor alpha chain - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 23-Jul-1999  
C:Accession: S35770  
R:Wedderburn, L.R.

Submitted to the EMBL Data Library, June 1993

A:Reference number: S35769

A:Accession: S35770

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-82 <WED>

A:Cross-references: EMBL:222966; NID:g312155; PIDN:CAA80539.1; PID:g312156  
C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: T-cell receptor

Query Match 60.9%; Score 39; DB 2; Length 82;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GSARQLTF 13  
Db 60 GSARQLTF 67

RESULT 2

AE2729

Cytochrome oxidase assembly factor cox15 [imported] - Agrobacterium tumefaciens (stra  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C:Accession: AE2729

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McCl  
Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam  
ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193

A:Accession: AE2729

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-356 <KUR>

A:Cross-references: GB:AE008688; PIDN:AAL42251.1; PID:gl7739647; GSPDB:GN00186  
A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: cox15

A:Map position: circular chromosome

Thu Jun 6 10:09:24 2002

Query Match 59.4%; Score 38; DB 1; Length 803;  
Best Local Similarity 70.0%; Pred. No. 37;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLVLGSGARQLT 10  
II:IIII:I  
DB 17 CLIVGSSAQO 26

RESULT 5

T09926  
cytidine deaminase (EC 3.5.4.5) CDA3 - Arabidopsis thaliana  
N:Alternate names: protein T16L4.140  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 11-Jan-2000  
C:Accession: T09926  
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K.  
submitted to the Protein Sequence Database, June 1999  
A:Reference number: Z16897  
A:Accession: T09926  
A:Molecule type: DNA  
A:Residues: 1-223 <BEV>  
A:Cross-references: EMBL:AL079344; GSPDB:GN00062; ATSP:T16L4.140  
A:Experimental source: cultivar Columbia; BAC clone T16L4  
C:Genetics:  
A:Gene: CDA3; ATSP:T16L4.140  
A:Map position: 4  
C:Superfamily: cdd protein  
C:Keywords: hydrolase

Query Match 57.8%; Score 37; DB 2; Length 223;  
Best Local Similarity 63.6%; Pred. No. 16;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CLVLGSGARQL 11  
IIIIII::  
DB 171 CLSLGSAEEI 181

RESULT 6

T09925  
cytidine deaminase (EC 3.5.4.5) CDA2 - Arabidopsis thaliana  
N:Alternate names: protein T16L4.130  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 03-Nov-2000  
C:Accession: T09925; T52131  
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K.  
submitted to the Protein Sequence Database, June 1999  
A:Reference number: Z16897  
A:Accession: T09925  
A:Molecule type: DNA  
A:Residues: 1-337 <BEV>  
A:Cross-references: EMBL:AL079344; GSPDB:GN00062; ATSP:T16L4.130  
A:Experimental source: cultivar Columbia; BAC clone T16L4  
R:Faivre-Nitschke, E.S.; Grienberger, J.M.; Gualberto, J.M.  
submitted to the EMBL Data Library April 1998  
A:Description: Cloning and characterisation of a cytidine deaminase gene family from  
A:Reference number: Z25972  
A:Accession: T52131  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-337 <FAI>  
A:Cross-references: EMBL:AJ005811; PIDN:CAA06710.1  
C:Genetics:  
A:Gene: CDA2; ATSP:T16L4.130; cda2  
A:Map position: 4  
C:Superfamily: cdd protein  
C:Keywords: hydrolase

Query Match 57.8%; Score 37; DB 2; Length 337;  
Best Local Similarity 63.6%; Pred. No. 24;

Query Match 59.4%; Score 38; DB 2; Length 356;  
Best Local Similarity 66.7%; Pred. No. 16;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLVLGSGARQLT 12  
IIIIII::II  
DB 28 CLVLGGATRLT 39

RESULT 3

F97510  
BA83fl1.2.1 (cox15 (yeast) homolog, cytochrome c oxidase assembly protein (isoform 1))  
C:Species: Agrobacterium tumefaciens  
C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C:Accession: F97510  
R:Goodner, B.; Wollam, C.; Allinger, D.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
A.; Liu, F.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A97359; PMID:11743194  
A:Accession: F97510  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-372 <KUR>  
A:Cross-references: GB:AE007869; PIDN:AAK87039.1; PID:gl15156289; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR\_C.2290  
A:Map position: circular chromosome

Query Match 59.4%; Score 38; DB 2; Length 372;  
Best Local Similarity 66.7%; Pred. No. 17;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLVLGSGARQLT 12  
IIIIII::II  
DB 44 CLVLVGGATRLT 55

RESULT 4

IJC83  
integrin, band 3 precursor - chicken  
N:Alternate names: CSAT antigen; JG22 antigen; RGD-receptor  
C:Species: Gallus gallus (chicken)  
C>Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 23-Jul-1999  
C:Accession: A23947  
R:Tamkun, J.W.; Desimone, D.W.; Fonda, D.; Patel, R.S.; Buck, C.; Horwitz, A.F.; Hynes,  
Cell 46, 271-282, 1986  
A:Title: Structure of integrin, a glycoprotein involved in the transmembrane linkage bet  
A:Reference number: A23947; MUID:86245073  
A:Accession: A23947  
A:Molecule type: mRNA  
A:Residues: 1-803 <VAM>  
A:Cross-references: GB:M14049; NID:g212213; PIDN:AAA48926.1; PID:g212214  
A:Experimental source: embryonic fibroblasts  
A:Note: the amino end of the mature protein is blocked  
C:Comment: This transmembrane complex may be the target of oncogenic transformation that  
skeleton. The cytoplasmic domain of this subunit contains a potential tyrosine-kinase pho  
C:Comment: The extracellular domain of this protein, like many membrane receptors, conta  
C:Superfamily: integrin beta chain; laminin-type EGF-like homology  
A:Keywords: cell adhesion; cytoskeleton; duplication; extracellular matrix; glycoprotein  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-803/Product: integrin, band 3 #status predicted <VAM>  
F:25-733/Domain: extracellular #status predicted <EXT>  
F:467-654/Region: cysteine-rich  
F:482-564,565-650/Region: duplication  
F:734-756/Domain: transmembrane #status predicted <MEM>  
F:757-803/Domain: intracellular #status predicted <INT>  
F:216,273,367,410,421,433,445,486,525,589,624,674/Binding site: carbohydrate (Asn) (cova  
F:788/Binding site: phosphate (Tyr) (covalent) #status predicted

^ Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CLVLSGSARQL 11  
||| ||||| :  
Db 186 CLTSLGSAGEI 196

RESULT 7

NKVGCU

core protein p20 - cucumber necrosis virus

C:Species: cucumber necrosis virus

A:Note: host Cucumis sativus (cucumber)

C:Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 23-Jul-1999

C:Accession: JA0133

R:Rochon, D.M.; Tremaine, J.H.

Virolgy 189, 251-259, 1989

A:Reference: Complete nucleotide sequence of the cucumber necrosis virus genome.

A:Accession: JA0133

A:Molecule type: genomic RNA

A:Residues: 1-173 <ROC>

C:Cross-references: GB:M25270; NID:G323338; PIDN:AAA42906.1; PID:G323343

C:Comment: The genome is a single-stranded, positive-sense RNA. It codes for 33K, 92K, 4

C:Superfamily: tombusvirus core protein p19

C:Keywords: core protein

Query Match 56.2%; Score 36; DB 1; Length 173;  
Best Local Similarity 63.6%; Pred. No. 19;

Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLVLSGSARQL 11  
||| ||| :  
Db 120 CLTSLGSARQL 130

RESULT 8

E84668

hypothetical protein At2g27080 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: E84668

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

Wu, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-260 <STO>

C:Cross-references: GB:AE002093; NID:G3085338; PIDN:AACT7866.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g27080

A:Map position: 2

Query Match 56.2%; Score 36; DB 2; Length 260;  
Best Local Similarity 81.8%; Pred. No. 29;

Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 LVLSGSARQL 12

||||| |||

Db 181 LVLSGSARQL 191

RESULT 9

F87249

enoyl-CoA hydratase/isomerase family protein [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001

C:Accession: F87249

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: F87249

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-262 <STO>

A:Cross-references: GB:AE005673; NID:G13421092; PIDN:AAK21394.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0006

C:Superfamily: naphthoate synthase; enoyl-CoA hydratase homology

Query Match 56.2%; Score 36; DB 2; Length 262;  
Best Local Similarity 66.7%; Pred. No. 29;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVLSGSAR 9

||||| :

Db 55 CIVLTGSAR 63

RESULT 10

S74441

iron(III) diclitate transport system permease protein fecB - Synechocystis sp. (strai

N:Alternate names: protein sir1319

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000

C:Accession: S74441

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,

O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys

S.

A:Reference number: S74322; MUID:97061201

A:Accession: S74441

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-315 <KAN>

A:Cross-references: EMBL:D90899; GB:AB001339; NID:G1651650; PIDN:BAAL6593.1; PID:G165

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Gene: fecB

C:Superfamily: iron(III) diclitate transport protein

C:Keywords: iron transport

Query Match 56.2%; Score 36; DB 1; Length 315;  
Best Local Similarity 63.6%; Pred. No. 36;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CLVLSGSARQL 11

||||| :

Db 11 CLVLFSGAKQV 21

RESULT 11

AB0961

Cytochrome c-type biogenesis protein H2 [imported] - Salmonella enterica subsp. enter

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001

C:Accession: AB0961

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

, S.; Moulie, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

Thu Jun 6 10:09:24 2002

A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AB0961  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-347 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD03182.1; PID:g16504815; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: ccmH2

Query Match 56.2%; Score 36; DB 2; Length 347;  
 Best Local Similarity 72.7%; Pred. No. 39;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 LVLSGSARQLT 12  
 ||:|||||  
 Db 14 LVISGSARATT 24

RESULT 12  
 AE0787  
 cytochrome c-type biogenesis protein H1 [imported] - *Salmonella enterica* subsp. *enterica*  
 C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi  
 A:Note: this species has also been called *Salmonella typhi*  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
 C:Accession: AE0787  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, J.; T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AE0787  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-347 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD07479.1; PID:g16503474; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: ccmH1

Query Match 56.2%; Score 36; DB 2; Length 347;  
 Best Local Similarity 72.7%; Pred. No. 39;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 LVLSGSARQLT 12  
 ||:|||||  
 Db 14 LVISGSARATT 24

RESULT 13  
 W2WLEP  
 E2 protein - European elk papillomavirus  
 C:Species: European elk papillomavirus  
 C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 11-May-2000  
 C:Accession: D29499; D94457; D94506  
 R:Rhola, H.; Bergman, P.; Stroem, A.C.; Moreno-Lopez, J.; Pettersson, U.  
 Gene 50, 195-205, 1986  
 A:Title: Organization and expression of the transforming region from the European elk pa  
 A:Reference number: A91567; PMID:87219878  
 A:Accession: D29499  
 A:Molecule type: DNA  
 A:Residues: 1-415 <AHO>  
 A:Cross-references: GB:M15953; NID:g333025; PIDN:AAA66854.1; PID:g484020  
 R:Eriksson, A.  
 unpublished results 1987, cited by GenBank  
 A:Reference number: A94457  
 A:Accession: D94457  
 A:Molecule type: DNA  
 A:Residues: 1-415 <BRI>  
 A:Cross-references: GB:M15953; NID:g333025; PIDN:AAA66854.1; PID:g484020

R:Pettersson, U.  
 submitted to GenBank, August 1987  
 A:Reference number: A94506  
 A:Accession: D94506  
 A:Molecule type: DNA  
 A:Residues: 1-415 <PET>  
 A:Cross-references: GB:M15953; NID:g333025; PIDN:AAA66854.1; PID:g484020  
 C:Superfamily: papillomavirus E2 protein  
 C:Keywords: early protein

Query Match 56.2%; Score 36; DB 1; Length 415;  
 Best Local Similarity 60.0%; Pred. No. 47;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CLVLGSARQ 10  
 ||:|||||  
 Db 333 CLILSGNGQ 342

RESULT 14  
 E87318  
 hypothetical protein CC0559 [imported] - *Caulobacter crescentus*  
 C:Species: *Caulobacter crescentus*  
 C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C:Accession: E87318  
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, R.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko, J.; Moltava, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A:Title: Complete genome sequence of *Caulobacter crescentus*.  
 A:Reference number: A87249; PMID:21173698; PMID:11259647  
 A:Accession: E87318  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-606 <STO>  
 A:Cross-references: GB:AE005673; NID:g13421751; PIDN:AAK22545.1; GSPDB:GN00148  
 C:Genetics:  
 A:Gene: CC0559

Query Match 56.2%; Score 36; DB 2; Length 606;  
 Best Local Similarity 58.3%; Pred. No. 69;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 CLVLGSARQLT 12  
 ||:|||||  
 Db 501 CLVADGQAKALT 512

RESULT 15  
 S36550  
 E6 protein - human papillomavirus type 3  
 C:Species: human papillomavirus type 3  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 26-Aug-1999  
 C:Accession: S36550  
 R:Belius, H.; Hofmann, B.  
 submitted to the EMBL Data Library, August 1993  
 A:Description: Primer-directed sequencing of human papillomavirus types.  
 A:Reference number: S36469  
 A:Accession: S36550  
 A:Molecule type: DNA  
 A:Residues: 1-152 <DEL>  
 A:Cross-references: EMBL:X74462; NID:g397005; PIDN:CAA52469.1; PID:g397006  
 C:Superfamily: papillomavirus E6 protein  
 C:Keywords: DNA binding; early protein; nucleus; zinc finger

Query Match 54.7%; Score 35; DB 2; Length 152;  
 Best Local Similarity 46.2%; Pred. No. 27;  
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 CLVLGSARQLTF 13

Thu Jun 6 10:09:24 2002

us-08-881-509-6.open.rpr

Page 5

Db       ||::| ||::| :  
69 CULVEGIARRLKY 81

Search completed: June 6, 2002, 06:16:22  
Job time: 66 sec

05



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:16:26 ; Search time 13.46 Seconds  
(without alignments)  
37.396 Million cell updates/sec

Title: US-08-881-509-6

Perfect score: 64

Sequence: 1 CLVLSGSARQLTF 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	59.4	803	1 ITBL_CHICK	P07228 gallus gall
2	36	56.2	173	1 VP19_CNV	P15184 cucumber ne
3	36	56.2	415	1 VE2_PAPVE	P11329 european el
4	35	54.7	152	1 VE6_HPV03	P36799 human papil
5	35	54.7	254	1 ATTA_TRINI	P50725 trichoplusi
6	35	54.7	336	1 G3P_CORGL	O01651 corynebacte
7	35	54.7	379	1 VAP1_TORCA	P19333 torpedos cal
8	35	54.7	905	1 CH02_SCHPO	O74787 schizosacch
9	34.5	53.9	2039	1 CCH1_YEAST	P50077 saccharomyc
10	34	53.1	148	1 VE6_HPV10	P36802 human papil
11	34	53.1	180	1 YCBW_ECOLI	P75862 escherichia
12	34	53.1	357	1 PMFE_PROMI	P53522 proteus mir
13	34	53.1	388	1 YGBK_ECOLI	O04019 arabidopsis
14	34	53.1	419	1 PRSA_ARATH	P23431 mesocricetu
15	34	53.1	422	1 ZP3_MESAU	P54776 lycopersico
16	34	53.1	423	1 PRSA_LYCES	O23894 brassica ca
17	34	53.1	424	1 PRSA_BRACH	P46465 oryza sativ
18	34	53.1	429	1 PRSA_ORYSA	O30994 rhizobium m
19	34	53.1	442	1 FTSA_RHIME	P02789 rattus norv
20	34	53.1	530	1 UDB8_RAT	O9uelp homo sapien
21	34	53.1	533	1 FT23_HUMAN	O74712 candida alb
22	34	53.1	838	1 HIS2_CANAL	O00744 homo sapien
23	33.5	52.3	389	1 WN1B_HUMAN	P48614 mus musculu
24	33.5	52.3	389	1 WN1B_MOUSE	P43907 haemophilus
25	33	51.6	193	1 Y983_HABIN	P10598 mus musculu
26	33	51.6	231	1 CASE_MOUSE	O9qzx7 mus musculu
27	33	51.6	339	1 SRR_MOUSE	O9qz24 homo sapien
28	33	51.6	340	1 SRR_MOUSE	P03123 deer papill
29	33	51.6	416	1 VE2_PAPVD	O30991 agrobacteri
30	33	51.6	443	1 FTSA_AGRTU	P77809 actinobacil
31	33	51.6	494	1 G6PD_ACTAC	P58229 escherichia
32	33	51.6	511	1 XASA_ECO57	O33013 mycobacteri
33	33	51.6	521	1 SR5A_MTCLE	

34	33	51.6	530	1 UDB8_RAT	P36511 rattus norv
35	33	51.6	610	1 RNFC_STRPL	P11220 streptomyce
36	33	51.6	740	1 RNFC_ECO57	P58324 escherichia
37	33	51.6	740	1 RNFC_ECOLI	P77611 escherichia
38	33	51.6	1235	1 DPOL_PYRHO	O59610 pyrococcus
39	33	51.6	2715	1 G156_PAPRI	P13837 paramycium
40	33	51.6	3396	1 POLG_DENIS	P33478 d genome po
41	32.5	50.8	1500	1 CP5M_HUMAN	P31327 homo sapien
42	32.5	50.8	1500	1 CP5M_HUMAN	P07756 rattus norv
43	32	50.0	131	1 YKAN_ECOLI	P45465 escherichia
44	32	50.0	146	1 VE6_HPV28	P50802 human papil
45	32	50.0	188	1 RUVC_MYCTU	Q50627 mycobacteri

## ALIGNMENTS

RESULT 1  
ID ITBL\_CHICK STANDARD; PRT; 803 AA.  
AC P07228;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DE Integrin beta-1 precursor (CSAT antigen) (JG22 antigen) (RGD-receptor).  
GN ITGB1.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
OC NCBI\_TaxID=9031;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryonic fibroblast;  
RX MEDLINE=86245073; PubMed=3487386;  
RA Tamkun J.W., Desimone D.W., Fonda D., Patel R.S., Buck C., Horwitz A.F., Hynes R.O.;  
RT "Structure of Integrin, a glycoprotein involved in the transmembrane linkage between fibronectin and actin.";  
RL Cell 46:271-282(1986).  
CC -!- FUNCTION: INTEGRINS ALPHA-1/BETA-1, ALPHA-2/BETA-1, ALPHA-10/BETA-1 AND ALPHA-11/BETA-1 ARE RECEPTORS FOR COLLAGEN. INTEGRINS ALPHA-1/BETA-1 AND ALPHA-2/BETA-2 RECOGNIZE THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-E-R IN COLLAGEN. INTEGRINS ALPHA-2/BETA-1, ALPHA-3/BETA-1, ALPHA-4/BETA-1, ALPHA-5/BETA-1, ALPHA-6/BETA-1, ALPHA-10/BETA-1, ALPHA-11/BETA-1 AND ALPHA-V/BETA-1 ARE RECEPTORS FOR FIBRONECTIN. ALPHA-4/BETA-1 AND ALPHA-V/BETA-1 ARE RECEPTORS FOR THE ALTERNATIVELY SPLICED CS-1 AND CS-5 REGIONS OF FIBRONECTIN. INTEGRIN ALPHA-5/BETA-1 IS A RECEPTOR FOR FIBRINOGEN. INTEGRIN ALPHA-1/BETA-1, ALPHA-2/BETA-1, ALPHA-6/BETA-1 AND ALPHA-7/BETA-1 ARE RECEPTORS FOR LAMININ. INTEGRIN ALPHA-4/BETA-1 IS A RECEPTOR FOR VCAM1 AND RECOGNIZES THE SEQUENCE Q-I-D-S IN VCAM1. INTEGRIN ALPHA-9/BETA-1 IS A RECEPTOR FOR VCAM1, CYTOTACTIN AND OSTEOPOINTIN. IT RECOGNIZES THE SEQUENCE A-E-I-D-G-I-E-L IN CYTOTACTIN. INTEGRIN ALPHA-3/BETA-1 IS A RECEPTOR FOR EPLIGRIN AND THROMBOSPONDIN. INTEGRIN ALPHA-V/BETA-1 IS A RECEPTOR FOR VITROTECTIN. BETA-1 INTEGRINS RECOGNIZE THE SEQUENCE R-G-D IN A WIDE ARRAY OF LIGANDS.  
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-1 ASSOCIATES WITH EITHER ALPHA-1, ALPHA-2, ALPHA-3, ALPHA-4, ALPHA-5, ALPHA-6, ALPHA-7, ALPHA-8, ALPHA-9, ALPHA-10, ALPHA-11 OR ALPHA-V.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.

-----  
This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch).

CC EMBL; M14049; AAA48926.1; -

DR PIR; A23947; IJCH3.

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR002369; Integrin\_B.

DR InterPro; IPR001169; Integrin\_beta\_c.

DR InterPro; IPR003659; PSI.

DR InterPro; IPR002035; VWFA.

DR Pfam; PF00362; Integrin\_B; 1.

DR PRINTS; PR01186; INTEGRINB.

DR ProDom; PD001811; Integrin\_B; 1.

DR SMART; SM00187; INB; 1.

DR SMART; SM00423; PSI; 1.

DR SMART; SM00327; VWFA; 1.

DR PROSITE; PS00243; INTEGRIN\_BETA; 3.

DR PROSITE; PS00022; EGF\_1; UNKNOWN.2.

DR Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein.

KW Repeat; Phosphorylation; Signal.

FT SIGNAL 1 24 BY SIMILARITY.

FT CHAIN 25 803 INTEGRIN BETA-1.

FT DOMAIN 25 733 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 734 756 POTENTIAL.

FT DOMAIN 757 803 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 144 382 VWFA-LIKE.

FT DOMAIN 471 520 4 CYSTEINE-RICH TANDEM REPEATS.

FT REPEAT 471 520 I.

FT REPEAT 521 564 II.

FT REPEAT 565 603 III.

FT REPEAT 604 640 IV.

FT DISULFID 31 469 BY SIMILARITY.

FT DISULFID 39 49 BY SIMILARITY.

FT DISULFID 42 79 BY SIMILARITY.

FT DISULFID 52 68 BY SIMILARITY.

FT DISULFID 211 217 BY SIMILARITY.

FT DISULFID 265 305 BY SIMILARITY.

FT DISULFID 405 419 BY SIMILARITY.

FT DISULFID 439 696 BY SIMILARITY.

FT DISULFID 467 471 BY SIMILARITY.

FT DISULFID 482 491 BY SIMILARITY.

FT DISULFID 491 530 BY SIMILARITY.

FT DISULFID 496 505 BY SIMILARITY.

FT DISULFID 507 521 BY SIMILARITY.

FT DISULFID 536 541 BY SIMILARITY.

FT DISULFID 538 573 BY SIMILARITY.

FT DISULFID 543 558 BY SIMILARITY.

FT DISULFID 560 565 BY SIMILARITY.

FT DISULFID 579 584 BY SIMILARITY.

FT DISULFID 581 612 BY SIMILARITY.

FT DISULFID 586 595 BY SIMILARITY.

FT DISULFID 597 604 BY SIMILARITY.

FT DISULFID 618 623 BY SIMILARITY.

FT DISULFID 620 666 BY SIMILARITY.

FT DISULFID 625 635 BY SIMILARITY.

FT DISULFID 638 641 BY SIMILARITY.

FT DISULFID 645 728 BY SIMILARITY.

FT DISULFID 651 728 BY SIMILARITY.

FT DISULFID 670 704 BY SIMILARITY.

FT MOD\_RES 25 25 BLOCKED.

FT MOD\_RES 788 788 PHOSPHORYLATION (BY TYR-KINASES).

FT CARBOHYD 216 216 (POTENTIAL).

FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 367 367 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 421 431 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 445 445 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 486 486 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 525 525 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 589 589 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 624 624 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 674 674 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 803 AA; 88553 MW; 2F6FEFCDF2C80457 CRC64;

Query Match 59.4%; Score 38; DB 1; Length 803;

Best Local Similarity 70.0%; Pred. No. 14; Gaps 0;

Matches 7; Conservative 2; Mismatches 1; Indels 0;

QY 1 CLVLGSGSARQ 10

DB 17 CLVLGSGSARQ 26

||: ||||: |

RESULT 2

VP19\_CNV STANDARD; PRT; 173 AA.

ID VP19\_CNV

AC P15184;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 01-AUG-1990 (Rel. 15, Last annotation update)

DE Core protein P19 (P20).

OS Cucurbit necrosis virus (CNV).

OC Cucurbit necrosis virus (CNV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;

OC Tombusvirus.

OC NCBI\_TaxID=12143;

OX [1]

RN SEQUENCE FROM N.A.

RP MEDLINE=89204896; PubMed=2705296;

RA Rochon D.M., Tremaine J.H.;

RT "Complete nucleotide sequence of the cucumber necrosis virus genome.";

RL Virology 169:251-259(1989).

CC -1- SIMILARITY: TO OTHER TOMBUSVIRUSES CORE PROTEIN P19.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; M25270; AAA42906.1; -

DR PIR; JAO133; NKVGGU.

DR Core protein.

KW SEQUENCE 173 AA; 19808 MW; D05F870D4AA4C80F CRC64;

Query Match 56.2%; Score 36; DB 1; Length 173;

Best Local Similarity 63.6%; Pred. No. 7;

Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLVLGSGSARQ 11

DB 120 CLVLGSGSARQ 130

||: ||||: |

RESULT 3

VE2\_PAPVE STANDARD; PRT; 415 AA.

ID VE2\_PAPVE

AC P11329;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Probable regulatory protein E2.

OS European elk papillomavirus (EEPV).

OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;

OC Papillomavirus.

OC NCBI\_TaxID=10565;

OX [1]

RN SEQUENCE FROM N.A.

RP MEDLINE=87219878; PubMed=3034730;

RA Ahola H., Bergman P., Stroem A.C., Moreno-Lopez J., Petterson U.;

RT "Organization and expression of the transforming region from the







SQ SEQUENCE 180 AA; 20589 MW; 7A8CC66ECE2D1B87 CRC64;

Query Match 53.1%; Score 34; DB 1; Length 180;  
Best Local Similarity 54.5%; Pred. No. 18; Indels 0; Gaps 0;  
Matches 6; Conservative 3; Mismatches 2;

QY 1 CLVLSSGARQLT 11  
Db 144 CVVIAGRAMQL 154

RESULT 12  
PMFE\_PROMI STANDARD; PRT; 357 AA.

AC P53522;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Putative minor fimbrial subunit pmfE precursor.  
GN PMFE.  
OS Proteus mirabilis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Proteus.  
OX NCBI\_TaxID=584;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HI4320;  
RX MEDLINE=95047519; PubMed=7959033;  
RA Massad G., Mobley H.L.T.;  
RT "Genetic organization and complete sequence of the Proteus mirabilis  
pmf fimbrial operon";  
RL Gene 150:101-104(1994).  
CC -1- SUBCELLULAR LOCATION: Fimbria.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z35428; CAA84592.1; -.  
KW Fimbria; Signal.  
FT SIGNAL 1 28 POTENTIAL.  
FT CHAIN 29 357 PUTATIVE MINOR FIMBRIAL SUBUNIT PMFE.  
SQ SEQUENCE 357 AA; 38876 MW; 59AD7E566D4899AA CRC64;

Query Match 53.1%; Score 34; DB 1; Length 357;  
Best Local Similarity 63.6%; Pred. No. 36;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 VLSSGARQLT 13  
Db 154 ISSGSSGQLT 164

RESULT 13  
YGBK\_ECOLI STANDARD; PRT; 388 AA.

AC Q46889;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein ygbK.  
GN YGBK OR B2737.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]

use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; X74465; CAA52489.1; -.  
DR PIR; S36532; S36532.  
DR InterPro; IPR001334; E6.  
DR Pfam; PF00518; E6; 1.  
KW Early protein, DNA-binding; Nuclear protein; Zinc-finger.  
FT ZN\_FING 29 65 POTENTIAL.  
FT ZN\_FING 102 138 POTENTIAL.  
SQ SEQUENCE 148 AA; 17563 MW; EFCAG68C51E61DB1A CRC64;

Query Match 53.1%; Score 34; DB 1; Length 148;  
Best Local Similarity 46.2%; Pred. No. 15;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 CLVLSSGARQLT 13  
Db 65 CLVLQGVRLKY 77

RESULT 11  
YCBW\_ECOLI STANDARD; PRT; 180 AA.

AC P75862;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein ycbW.  
GN YCBW OR B0946.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
PT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
RA Sanpei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horiuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 12.7-28.0 min region on the linkage map.";  
RL DNA Res. 3:137-155(1996).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF000196; AAC74032.1; ALT\_INIT.  
DR EMBL; D90732; BAA35701.1; ALT\_INIT.  
DR EcoGene; EG13715; ycbW.  
KW Hypothetical protein; Complete proteome.

```

RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RA MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick J.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -1- SIMILARITY: STRONG, TO H. INFLUENZAE H1N1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U29579; AAA69247.1; -
DR EMBL; AE000357; AAC75779.1; -
DR EcoGene; EG13105; ygbK.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 388 AA; 41339 MW; 5824A120E9393892 CRC64;

Query Match 53.1%; Score 34; DB 1; Length 388;
Best Local Similarity 63.6%; Pred. No. 40;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 LVLSGSRQUT 12
Db :||||| 1:1
253 VVLGSGSQMT 263

RESULT 14
PRSA_ARATH
ID PRSA_ARATH STANDARD; PRT; 419 AA.
AC 004019;
AT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 26S protease regulatory subunit 6A homolog (TAT-binding protein
DE homolog 1) (TBP-1).
GN AT1G09100 OR F7G19.2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-CV. COLUMBIA;
RA MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizier L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RL Nature 408:816-820(2000).

```

CC -1- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE) COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE 26S COMPLEX (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (Potential).

CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AC000106; AAB70397.1; -

DR InterPro; IPR003593; AAA.

DR InterPro; IPR003960; AAA\_sub.

DR InterPro; IPR003959; AAA\_subfam.

DR Pfam; PF00004; AAA; 1.

DR SMART; SM00382; AAA; 1.

DR PROSITE; PS00674; AAA; 1.

KW Proteasome; ATP-binding; Nuclear protein.

FT NP\_BIND 207 214

SQ SEQUENCE 419 AA; 46663 MW; BD2113676BCE44A CRC64;

Query Match 53.1%; Score 34; DB 1; Length 419;

Best Local Similarity 53.8%; Pred. No. 43;

Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 CLVLGSGARQLTF 13

Db :||| 1:1

101 CVLKTSTQRTIF 113

RESULT 15

ZP3\_MESAU

ID ZP3\_MESAU STANDARD; PRT; 422 AA.

AC P23491;

AT 01-NOV-1991 (Rel. 20, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DE 30-MAY-2000 (Rel. 39, Last annotation update)

DE Zon pellucida sperm-binding protein 3 precursor (Zon pellucida glycoprotein zp3) (Sperm receptor) (Zon pellucida protein C).

GN ZP3.

OS Mesocricetus auratus (Golden hamster).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.

OC NCBI\_TaxID=10036;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Ovary;

RX MEDLINE=91078540; PubMed=2257975;

RX Kinloch R.A., Ruiz-Seller B., Wasserman P.M.;

RT "Genomic organization and polypeptide primary structure of zon pellucida glycoprotein hzp3, the hamster sperm receptor.";

RT Dev. Biol. 142:414-421(1990).

CC -1- FUNCTION: FUNCTIONS AS A SPERM-RECEPTOR. IT IS RESPONSIBLE FOR SPERM-ADHESION TO THE ZON PELLUCIDA, AND MAY CONTRIBUTE TO THE SPECIES-SPECIFICITY OF THE INSEMINATION.

CC -1- SUBUNIT: ZP3 FORMS WITH ZP1 AND ZP2 THE ZON PELLUCIDA, IN WHICH ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Extracellular matrix.

CC -1- TISSUE SPECIFICITY: OOCYTES.

CC -1- DEVELOPMENTAL STAGE: GROWING OOCYTES.

CC -1- PTM: SULFATED GLYCOPROTEIN WITH O-LINKED OLIGOSACCHARIDES.

CC -1- SIMILARITY: CONTAINS 1 ZP DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

Thu Jun 6 10:09:24 2002

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: M63629; AAA37079.1; .  
DR InterPro: IPR001507; zona\_pellucida.  
DR Pfam: PF00100; zona\_pellucida; 1.  
DR PRINTS: PR00023; ZPELUCIDA.  
DR SMART: SM00241; ZP; 1.  
DR PROSITE: PS00682; ZP\_DOMAIN; 1.  
KW Glycoprotein; Signal; Sulfation; Sperm; Receptor; Transmembrane;  
KW Extracellular matrix.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 422 ZONA PELLUCIDA SPERM-BINDING PROTEIN 3.  
FT DOMAIN 23 386 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 387 407 POTENTIAL.  
FT DOMAIN 408 422 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 45 306 ZP.  
FT DOMAIN 119 158 PRO-RICH.  
FT DOMAIN 208 257 PRO-RICH.  
FT CARBOHYD 146 146 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 422 AA; 45827 MW; DQF95BE7FF8E7E01 CRC64;

Query Match 53.1%; Score 34; DB 1; Length 422;  
Best Local Similarity 60.0%; Pred. No. 43;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 CLVLSGSARQ 10  
||| | | | |  
Db 10 CLLLCGGAKQ 19

Search completed: June 6, 2002, 06:22:27  
Job time: 361 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:15:51 ; Search time 41.63 Seconds  
(without alignments)  
54.022 Million cell updates/sec

Title: US-08-881-509-6

Perfect score: 64

Sequence: 1 CLVLSSARQLTF 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_protist.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	59.4	222	2 Q9F598	Q9f598 agrobacteri
2	37	57.8	223	10 Q9S847	Q9s847 arabidopsis
3	37	57.8	322	10 Q9T333	Q9t333 arabidopsis
4	37	57.8	337	10 Q65571	Q65571 arabidopsis
5	37	57.8	337	10 Q9SWZ4	Q9swz4 arabidopsis
6	37	57.8	552	12 Q91SH2	Q91sh2 guinea pig
7	36	56.2	173	12 Q9IES5	Q9ies5 cucumber ne
8	36	56.2	173	12 Q9IES9	Q9ies9 cucumber ne
9	36	56.2	173	12 Q9IES8	Q9ies8 cucumber ne
10	36	56.2	173	12 Q9IES7	Q9ies7 cucumber ne
11	36	56.2	173	12 Q9IES6	Q9ies6 cucumber ne
12	36	56.2	243	10 Q9ATM2	Q9atm2 zea mays (m
13	36	56.2	260	10 Q9ZVD2	Q9zvd2 arabidopsis
14	36	56.2	262	16 Q9AC55	Q9ac55 caulobacter
15	36	56.2	315	16 P72593	P72593 synechocyst
16	36	56.2	606	16 Q9AAN8	Q9aan8 caulobacter

## ALIGNMENTS

RESULT 1

Q9F598 PRELIMINARY; PRT; 222 AA.  
AC Q9F598;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE RIORF157 PROTEIN.  
GN RIORF157.  
OS Agrobacterium rhizogenes.  
OG Plasmid pR11724.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=359,  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF03-01724;  
RA Moriguchi K., Maeda Y., Satou M., Satuti N., Kataoka M., Tanaka N.,  
RA Yoshida K.;  
RT "The complete nucleotide sequence of a Ri (root inducing) plasmid  
RT indicates its chimerical structure between Ti and Sym plasmids.";  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF03-01724;  
RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;  
RT "Analysis of unique variable region of a plant root inducing plasmid,  
RT pR11724, by the construction of its physical map and library.";  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF03-01724;  
RA Moriguchi K., Nishida T., Maeda Y., Tanaka N., Yoshida K.;  
RT "Genome structure of Ri plasmid (1): Construction of linking library  
RT and physical map of pR11724 in Japanese Agrobacterium.";  
RL Nucleic Acids Symp. Ser. 39:189-190(1998).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF03-01724;

Q98ar4 rhizobium 1  
Q9ahx8 pseudomonas  
Q9ilc7 white spot  
Q918h9 white spot  
Q9su87 arabidopsis  
Q9syu5 arabidopsis  
Q83110 mouse adeno  
Q96381 hypanthia  
Q9zt35 arabidopsis  
Q9hgk3 debaromyce  
Q95sz9 drosophila  
P91632 drosophila  
Q9f839 micromonospor  
Q9ind9 cactus viru  
Q9veg7 drosophila  
Q61603 drosophila  
Q9blv5 leishmania  
Q962d5 drosophila  
Q9su10 arabidopsis  
Q03885 saccharomyc  
P70799 azotobacter  
Q02489 saccharomyc  
Q9kjb0 flavobacter  
Q9kia8 flavobacter  
Q9kia7 flavobacter  
Q31435 bacillus su  
P94526 bacillus su  
Q93s33 prevotella  
Q05073 bacillus ps

17 35 54.7 199 16 Q98AR4  
18 35 54.7 257 2 Q9AHX8  
19 35 54.7 292 12 Q91LC7  
20 35 54.7 292 12 Q918H9  
21 35 54.7 307 10 Q9SU87  
22 35 54.7 308 10 Q9SYU5  
23 35 54.7 316 12 Q83110  
24 35 54.7 320 5 Q96381  
25 35 54.7 329 10 Q9ZT35  
26 35 54.7 357 3 Q9HGK3  
27 35 54.7 364 5 Q95SZ9  
28 35 54.7 386 5 Q91632  
29 35 54.7 436 2 Q9F839  
30 35 54.7 436 2 Q9F839  
31 35 54.7 1543 12 Q9IND9  
32 35 54.7 2703 5 Q9VEG7  
33 35 54.7 2715 5 Q61603  
34 35 54.7 3187 5 Q9BLV5  
35 54.7 147 5 Q962D5  
36 34 53.1 183 10 Q9SUL0  
37 34 53.1 199 3 Q03885  
38 34 53.1 217 2 P70799  
39 34 53.1 227 3 Q02489  
40 34 53.1 248 2 Q9KJB0  
41 34 53.1 248 2 Q9KJA8  
42 34 53.1 248 2 Q9KJA7  
43 34 53.1 256 16 Q31435  
44 34 53.1 272 16 P94526  
45 34 53.1 281 2 Q93S33  
301 2 Q05073



```

RN [3]
RP SEQUENCE FROM N.A.
RA Bevan M., Rose M., Hempel S., Entian K.-D., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Rose M., Hempel S., Entian K.-D., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ005811; CAA06710.1; -
DR EMBL; AF121877; AAD30445.1; -
DR EMBL; AL079344; CAB45322.1; -
DR EMBL; AL161575; CAB79720.1; -
DR HSSP; P13652; 1CWT
DR InterPro; IPR002125; dCMP_cyt_deam.
DR Pfam; PF00383; dCMP_cyt_deam; 1.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
KW Hydrolase.
SQ SEQUENCE 337 AA; 37106 MW; 74B8CFF113EB46FB CRC64;

Query Match 57.8%; Score 37; DB 10; Length 337;
Best Local Similarity 63.6%; Pred. No. 61;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CLVLGSGARQL 11
|| ||||| ::
Db 186 CLTSGSAGEI 196

RESULT 5
Q9SWZ4 PRELIMINARY; PRT; 337 AA.
AC Q9SWZ4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CYTIDINE DEAMINASE 2 (EC 3.5.4.5).
GN CDA2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eutrosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Faivre-Nitschke S.E., Grienbenberger J.M., Gualberto J.M.;
RT "Cloning and characterisation of a cytidine deaminase gene family from
RT Arabidopsis thaliana.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF080676; AAC69567.2; -
DR HSSP; P13652; 1CWT.
DR InterPro; IPR002125; dCMP_cyt_deam.
DR Pfam; PF00383; dCMP_cyt_deam; 1.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
KW Hydrolase.
SQ SEQUENCE 337 AA; 37110 MW; 64ECCFF113EF46F8 CRC64;

Query Match 57.8%; Score 37; DB 10; Length 337;
Best Local Similarity 63.6%; Pred. No. 61;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CLVLGSGARQL 11
|| ||||| ::

```

```

Db 186 CLTSGSAGEI 196

RESULT 6
Q9LSH2 PRELIMINARY; PRT; 552 AA.
AC Q9LSH2;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GP35.
OS Guinea pig cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Muromegalovirus.
OX NCBI_TaxID=33706;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu Y., Biegalka B.J.;
RT "Identification of a cluster of late genes in guinea pig
RT cytomegalovirus.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF355272; AAK43593.1; -
SQ SEQUENCE 552 AA; 63280 MW; 268404E1641D7C60 CRC64;

Query Match 57.8%; Score 37; DB 12; Length 552;
Best Local Similarity 54.5%; Pred. No. 1e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLVLGSGARQL 11
|| || || || || ::
Db 442 CLMEGAAQQL 452

RESULT 7
Q9IES5 PRELIMINARY; PRT; 173 AA.
AC Q9IES5;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE 19 KDA PROTEIN (FRAGMENT).
GN P19.
OS Cucumber necrosis virus (CNV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;
OC Tombusvirus.
OX NCBI_TaxID=12143;
RN [1]
RP SEQUENCE FROM N.A.
RA Obermeier C., Sears J.L., Liu H.Y., Schluster K.O., Ryder E.J.,
RA Duffus J.E., Koike S.T., Wisler G.C.;
RT "Characterization of distinct tombusviruses that cause diseases of
RT lettuce and tomato in the Western United States.";
RL Phytopathology 91:797-806(2001).
DR EMBL; AJ288927; CAC01093.1; -
FT NON_TER
FT SEQUENCE 173 AA; 19474 MW; 69244B482B45AFA9 CRC64;

Query Match 56.2%; Score 36; DB 12; Length 173;
Best Local Similarity 63.6%; Pred. No. 47;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLVLGSGARQL 11
|| || || || || ::
Db 120 CLTSGSRTLL 130

RESULT 8
Q9IES9 PRELIMINARY; PRT; 173 AA.
ID Q9IES9

```

AC Q9IES9: 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE 19 KDA PROTEIN (FRAGMENT).  
GN P19.  
OS Cucumber necrosis virus (CNV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;  
OC Tombusvirus.  
OX NCBI\_TaxID=12143;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LETTUCE;  
RA Obermeier C., Sears J.L., Liu H.Y., Schlueter K.O., Ryder E.J.,  
RA Duffus J.E., Koike S.T., Wisler G.C.;  
RT "Characterization of distinct tombusviruses that cause diseases of  
RT lettuce and tomato in the Western United States.";  
RL Phytopathology 91:797-806(2001).  
DR EMBL; AJ288921; CAC01091.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 173 AA; 19474 MW; 69244B482E45AFA9 CRC64;

Query Match 56.2%; Score 36; DB 12; Length 173;  
Best Local Similarity 63.6%; Pred. No. 47;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLVLSGSARQL 11  
DB 120 CLTSLGGSRRL 130

RESULT 11  
Q9IES6 PRELIMINARY; PRT; 173 AA.  
AC Q9IES6;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE 19 KDA PROTEIN (FRAGMENT).  
GN P19.  
OS Cucumber necrosis virus (CNV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;  
OC Tombusvirus.  
OX NCBI\_TaxID=12143;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LETTUCE;  
RA Obermeier C., Sears J.L., Liu H.Y., Schlueter K.O., Ryder E.J.,  
RA Duffus J.E., Koike S.T., Wisler G.C.;  
RT "Characterization of distinct tombusviruses that cause diseases of  
RT lettuce and tomato in the Western United States.";  
RL Phytopathology 91:797-806(2001).  
DR EMBL; AJ288922; CAC01092.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 173 AA; 19484 MW; 69244B482D8AFA9 CRC64;

Query Match 56.2%; Score 36; DB 12; Length 173;  
Best Local Similarity 63.6%; Pred. No. 47;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLVLSGSARQL 11  
DB 120 CLTSLGGSRRL 130

RESULT 12  
Q9ATM2 PRELIMINARY; PRT; 243 AA.  
AC Q9ATM2;  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)  
DE SMALL BASIC MEMBRANE INTEGRAL PROTEIN ZMSIP1-2.  
OS Zea mays (Maize).

AC Q9IES9: 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE 19 KDA PROTEIN (FRAGMENT).  
GN P19.  
OS Cucumber necrosis virus (CNV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;  
OC Tombusvirus.  
OX NCBI\_TaxID=12143;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LETTUCE;  
RA Obermeier C., Sears J.L., Liu H.Y., Schlueter K.O., Ryder E.J.,  
RA Duffus J.E., Koike S.T., Wisler G.C.;  
RT "Characterization of distinct tombusviruses that cause diseases of  
RT lettuce and tomato in the Western United States.";  
RL Phytopathology 91:797-806(2001).  
DR EMBL; AJ288919; CAC01089.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 173 AA; 19484 MW; 69244B482D8AFA9 CRC64;

Query Match 56.2%; Score 36; DB 12; Length 173;  
Best Local Similarity 63.6%; Pred. No. 47;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLVLSGSARQL 11  
DB 120 CLTSLGGSRRL 130

RESULT 9  
Q9IES8 PRELIMINARY; PRT; 173 AA.  
AC Q9IES8;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE 19 KDA PROTEIN (FRAGMENT).  
GN P19.  
OS Cucumber necrosis virus (CNV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;  
OC Tombusvirus.  
OX NCBI\_TaxID=12143;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LETTUCE;  
RA Obermeier C., Sears J.L., Liu H.Y., Schlueter K.O., Ryder E.J.,  
RA Duffus J.E., Koike S.T., Wisler G.C.;  
RT "Characterization of distinct tombusviruses that cause diseases of  
RT lettuce and tomato in the Western United States.";  
RL Phytopathology 91:797-806(2001).  
DR EMBL; AJ288920; CAC01090.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 173 AA; 19484 MW; 69244B482D8AFA9 CRC64;

Query Match 56.2%; Score 36; DB 12; Length 173;  
Best Local Similarity 63.6%; Pred. No. 47;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLVLSGSARQL 11  
DB 120 CLTSLGGSRRL 130

RESULT 10  
Q9IES7 PRELIMINARY; PRT; 173 AA.  
AC Q9IES7;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC Panicoidae; Andropogoneae; Zea.  
 RN NCBI\_TaxID=4577;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21140306; PubMed=11244102;  
 RA Chaumont F., Barrieu F., Wojcik E., Chrispeels M.J., Jung R.;  
 RT "Aquaporins constitute a large and highly divergent protein family in  
 RL Plant Physiol. 125:1206-1215(2001).  
 DR EMBL; AF326498; AAK26765.1; -;  
 DR InterPro: IPR000425; MIP.  
 DR Pfam: PF00230; MIP; 1.  
 DR PRINTS: PR00783; MINTRINSCP.  
 SQ SEQUENCE 243 AA; 25703 MW; 5B37A379977C8C9E CRC64;

Query Match 56.2%; Score 36; DB 10; Length 243;  
 Best Local Similarity 87.5%; Pred. No. 67;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVLSGSA 8  
 DB 175 CLVLSGAA 182  
 ID Q92VD2 PRELIMINARY; PRT; 260 AA.  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE AT2G27080 PROTEIN (AT2G27080/T20P8.13) (UNKNOWN PROTEIN).  
 GN AT2G27080.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLOMBIA;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
 RA Buel C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana.";  
 RL Nature 402:761-768(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLOMBIA;  
 RA Lin X.;  
 RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,  
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.;  
 RT "Arabidopsis cDNA clones";  
 RN Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]

RP SEQUENCE FROM N.A.  
 RA Lam B., Southwick A., Karlin-Neumann G., Nguyen M., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Chen H., Cheuk R.,  
 RA Chung M.K., Kim C., Lin J., Liu S.X., Pham P.K., Sakano H., Shinn P.,  
 RA Yamada K., Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC005623; AAC77866.1; -;  
 DR EMBL; AY052291; AAK96484.1; -;  
 DR EMBL; AF370591; AAK43910.1; -;  
 SQ SEQUENCE 260 AA; 28975 MW; 36872A3E08F24371 CRC64;

Query Match 56.2%; Score 36; DB 10; Length 260;  
 Best Local Similarity 81.8%; Pred. No. 72;  
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 LVLGSGARQLT 12  
 DB 181 LVLGSGSKIQT 191  
 ID Q9AC55 PRELIMINARY; PRT; 262 AA.  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE ENOYL-COA HYDRATASE/ISOMERASE FAMILY PROTEIN.  
 GN CC0006.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter.  
 OX NCBI\_TaxID=69394;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT "Complete genome sequence of Caulobacter crescentus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL; AE005675; AAK21994.1; -;  
 DR HSSP; P14604; 2DUB.  
 DR TIGR; CC0006;  
 DR InterPro: IPR001753; Enoyl-CoA\_hydrase.  
 DR Pfam; PF00378; ECH; 1.  
 DR PROSITE; PS00166; ENOYL-COA-HYDRATASE; 1.  
 KW Isomerase; Complete proteome.  
 SQ SEQUENCE 262 AA; 27631 MW; 9CB076DD3F6B9D71 CRC64;

Query Match 56.2%; Score 36; DB 16; Length 262;  
 Best Local Similarity 66.7%; Pred. No. 73;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVLSGSA 9  
 DB 55 CIVLTGSAK 63  
 ID P72593 PRELIMINARY; PRT; 315 AA.  
 AC P72593;  
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE	IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN FECB.
FE	FECB OR SLR1319.
OS	Synechocystis sp. (strain PCC 6803).
OC	Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX	NCBI_TaxID=1148;
NR	[1]
NR	SEQUENCE FROM N.A.
NR	MEDLINE=97061201; PubMed=8905231;
RR	Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA	Maye T., Hiroseawa M., Sugihara M., Sasamoto S., Kimura T.,
RA	Yajima N., Hironaka M., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA	Hosouchi T., Matsuno A., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA	Shimo S., Takeuchi C., Wada T.,
RA	Tabata S.;
RA	"Sequence analysis of the genome of the unicellular cyanobacterium
RT	Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT	entire genome and assignment of potential protein-coding regions.;"
RL	DNA Res. 3:109-136(1996).
RL	EMBL; D90899; BAA16593.1; -;
DR	InterPro; IPR002491; Peripla.BP.
DR	Pfam; PF01497; Peripla_BP_2; 1.
DR	complete proteome.
DR	SEQUENCE 315 AA; 34868 MW; 3F9480C26352A41 CRC64;
SO	

Query Match	56.2%	Score 36;	DB 16;	Length 315;
Best Local Similarity	63.6%;	pred. No. 88;		
Metabase 7. Conservative	2;	Mismatches 2;	Indels 0;	Gaps 0;

Qy 1 CLVLSGSARQL 11  
 ||||| | | :  
 Db 11 CLVLFSGCAKQV 21

Search completed: June 6, 2002, 06:22:08  
Job time: 377 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:15:16 ; Search time 52.04 Seconds  
(without alignments)  
27.747 Million cell updates/sec

Title: US-08-881-509-6

Perfect score: 64

Sequence: 1 CLVLGSGARQLTF 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

- 1: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	64	100.0	13	19	AAW47590
2	45.5	71.1	267	19	AAW47588
3	42.5	66.4	12	19	AAW47591
4	38	59.4	205	21	AAB56211
5	38	59.4	222	21	AAB56352
6	38	59.4	537	22	ABG26118
7	38	59.4	663	22	ABG08826
8	38	59.4	663	22	ABG29486
9	36	56.2	34	22	ABB32436
10	36	56.2	34	22	AAW58353
11	36	56.2	34	22	AAW18673

12	54.7	147	22	AAU39475	Propionibacterium
13	54.7	284	22	AG84910	Shrimp white spot
14	54.7	334	22	AG91501	C glutamicum prote
15	54.7	334	22	AB79276	Corynebacterium g1
16	54.7	386	22	AB83279	Drosophila melanog
17	54.7	436	22	AB82203	Megalomicin biosyn
18	54.7	2703	22	AB83299	Drosophila melanog
19	54.7	28	16	AA80130	TCR CDR3 V-alpha 1
20	53.1	44	22	AAW69627	Human bone marrow
21	53.1	44	22	AAW17440	Peptide #3874 enco
22	53.1	55	22	AB338722	Peptide #6228 enco
23	53.1	55	22	AAW59355	Human brain expres
24	53.1	55	22	AAW71910	Human bone marrow
25	53.1	55	22	AAW32191	Peptide #6228 enco
26	53.1	59	20	AAV13021	Human secreted pro
27	53.1	61	22	AAU53825	Propionibacterium
28	53.1	78	20	AAV36016	Extended human sec
29	53.1	99	22	AAU14591	Human novel protei
30	53.1	106	11	AAW06641	Histidinol dehydro
31	53.1	112	22	AAW74229	Human colon cancer
32	53.1	118	22	AAU14355	Human novel protei
33	53.1	119	21	AAU18823	Zea mays protein f
34	53.1	157	21	AAU18821	Zea mays protein f
35	53.1	174	22	AAU59698	Propionibacterium
36	53.1	182	21	AAW24465	Arabidopsis thalia
37	53.1	183	21	AAW24465	Arabidopsis thalia
38	53.1	183	21	AAW24465	Arabidopsis thalia
39	53.1	185	21	AAW24463	Arabidopsis thalia
40	53.1	185	21	AAW24463	Arabidopsis thalia
41	53.1	189	21	AAW24463	Arabidopsis thalia
42	53.1	269	19	AAW53952	Arabidopsis thalia
43	53.1	271	21	AAW40344	Bacillus subtilis
44	53.1	297	21	AAW76748	Human OREF ORF108
45	53.1	297	22	AAW06206	Human protein kina

## ALIGNMENTS

RESULT	1
AAW47590	
ID	AAW47590 standard; peptide; 13 AA.
XX	
AC	AAW47590;
XX	
DT	26-JUN-1998 (first entry)
XX	
DE	T-cell receptor CDR3 alpha-region.
XX	
KW	Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;
KW	prevention; therapy; tumour disease; renal cell carcinoma;
KW	CDR3.
XX	
OS	Homo sapiens.
XX	
PN	DE19625191-A1.
XX	
PD	02-JAN-1998.
XX	
PF	24-JUN-1996; 96DE-1025191.
XX	
PR	24-JUN-1996; 96DE-1025191.
XX	
PA	(BOEF ) BOEHRINGER MANNHEIM GMBH.
XX	
PI	Schendel D;
XX	
DR	WPI; 1998-053442/06.
DR	N-PSDB; AAV18707.
XX	
PT	Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT	diagnosis or therapy, especially of renal cell carcinoma
XX	

PS Example 1; Page 17; 30pp; German.

XX The present sequence is the CDR3 alpha-region of a human  
CC T-cell receptor (TCR), which can be used in the diagnosis,  
CC monitoring, prevention and therapy of a tumour disease,  
CC specifically renal cell carcinoma.

XX Sequence 13 AA;

Query Match 100.0%; Score 64; DB 19; Length 13;  
Best Local Similarity 100.0%; Pred. No. 6.2e-06;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVLGSGARQLTF 13  
||| ||||| |||||  
Db 1 clvlgsgarqltf 13

RESULT 2

AAW47588  
ID AAW47588 standard; Protein; 267 AA.

XX AAW47588;

XX 26-JUN-1998 (first entry)

XX T-cell receptor alpha-chain.

XX Alpha-chain; human; T-cell receptor; TCR; diagnosis; monitoring;  
KW prevention; therapy; tumour disease; renal cell carcinoma.

XX Homo sapiens.

XX DE19625191-A1.

XX 02-JAN-1998.

XX 24-JUN-1996; 96DE-1025191.

XX 24-JUN-1996; 96DE-1025191.

XX (BOEF ) BOEHRINGER MANNHEIM GMBH.

XX Schendel D;

XX WPI; 1998-053442/06.

XX N-PSDB; AAV18705.

XX Human T-cell receptor nucleic acids and poly:peptide(s) - for  
PT diagnosis or therapy, especially of renal cell carcinoma

XX Example 1; Pages 11-13; 30pp; German.

XX The present sequence is the alpha-chain of a human T-cell  
CC receptor (TCR), which can be used in the diagnosis, monitoring,  
CC prevention and therapy of a tumour disease, specifically renal  
CC cell carcinoma.

XX Sequence 267 AA;

Query Match 71.1%; Score 45.5; DB 19; Length 267;  
Best Local Similarity 84.6%; Pred. No. 0.93;  
Matches 11; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 CLVLGSGARQLTF 13

||| ||||| |||||  
Db 105 clv-gsgarqltf 116

RESULT 3

AAW47591

ID AAW47591 standard; peptide; 12 AA.

XX AAW47591;

XX 26-JUN-1998 (first entry)

XX T-cell receptor CDR3 alpha-region.

XX Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;  
KW prevention; therapy; tumour disease; renal cell carcinoma;  
KW CDR3.

XX Homo sapiens.

XX DE19625191-A1.

XX 02-JAN-1998.

XX 24-JUN-1996; 96DE-1025191.

XX 24-JUN-1996; 96DE-1025191.

XX (BOEF ) BOEHRINGER MANNHEIM GMBH.

XX Schendel D;

XX WPI; 1998-053442/06.

XX N-PSDB; AAV18708.

XX Human T-cell receptor nucleic acids and poly:peptide(s) - for  
PT diagnosis or therapy, especially of renal cell carcinoma

XX Example 1; Page 17; 30pp; German.

XX The present sequence is the CDR3 alpha-region of a human  
CC T-cell receptor (TCR), which can be used in the diagnosis,  
CC monitoring, prevention and therapy of a tumour disease,  
CC specifically renal cell carcinoma.

XX Sequence 12 AA;

Query Match 66.4%; Score 42.5; DB 19; Length 12;

Best Local Similarity 76.9%; Pred. No. 0.1;

Matches 10; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 1 CLVLGSGARQLTF 13

||| ||||| |||||  
Db 1 clv-tgsarqltf 12

RESULT 4

AAW47591

ID AAW47591 standard; Protein; 205 AA.

XX AAW47591;

XX 13-MAR-2001 (first entry)

XX Human secreted protein sequence encoded by gene 135 SEQ ID NO:305.

XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;  
KW fungicide; ophthalmological; gene therapy; pathological condition;  
KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;  
KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;  
KW cerebrovascular disorder; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; skin aging; food additive; preservative.

XX Homo sapiens.

XX



PN W0200070042-A1.  
 XX 23-NOV-2000.  
 PD 11-MAY-2000; 2000WO-US12788.  
 XX 13-MAY-1999; 99US-0134068.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;  
 PI Duan RD, Florence KA, Soppet DR;  
 PI WPI; 2000-679828/66.  
 DR N-PSDB; AAC99952.  
 XX Isolated nucleic acid molecule encoding a human secreted protein is  
 PT used in preventing, treating or ameliorating a medical condition -  
 PT Claim 11; Page 995-996; 1065pp; English.  
 XX The polynucleotide sequences given in AAC99818 to AAC99977 encode the  
 CC human secreted proteins given in AAB56077 to AAB56362. Human secreted  
 CC proteins have activities based on the tissues and cells the genes are  
 CC expressed in. Examples of activities include: immunosuppressive;  
 CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;  
 CC vasotropic; cerebroprotective; neutrotropic; neuroprotective; antibacterial;  
 CC virucide; fungicide; and ophthalmological. The human secreted  
 CC polynucleotides and proteins can be used to prevent, treat or ameliorate  
 CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,  
 CC dogs, chickens or sheep. They are also used in diagnosing a pathological  
 CC condition or susceptibility to a pathological condition. Disorders which  
 CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid  
 CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or  
 CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular  
 CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders  
 CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and  
 CC fungi and ocular disorders e.g. corneal infection. The proteins can also  
 CC be used to aid wound healing and epithelial cell proliferation, to  
 CC transplanted, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The proteins can also be used as a  
 CC food additive or preservative to increase or decrease storage  
 CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used  
 CC in the exemplification of the present invention.  
 XX Sequence 205 AA;  
 SQ  
 Query Match 59.4%; Score 38; DB 21; Length 205;  
 Best Local Similarity 53.8%; Pred. NO. 21;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 CLVLGSGARQLTF 13  
 ||| ||| :|  
 Db 88 cllcgssrttsf 100  
 RESULT 5  
 AAB56352  
 ID AAB56352 standard; Protein; 222 AA.  
 XX AAB56352;  
 XX 13-MAR-2001 (first entry)  
 DT Human secreted protein sequence encoded by gene 135 SEQ ID NO:446.  
 DE Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
 XX antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
 KW cerebroprotective; neutrotropic; neuroprotective; antibacterial; virucide;  
 KW fungicide; ophthalmological; gene therapy; pathological condition;  
 KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;

KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;  
 KW cerebrovascular disorder; angiogenesis; nervous system disorder;  
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
 KW wound healing; skin aging; food additive; preservative.  
 XX Homo sapiens.  
 OS W0200070042-A1.  
 XX 23-NOV-2000.  
 PD 11-MAY-2000; 2000WO-US12788.  
 XX 13-MAY-1999; 99US-0134068.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;  
 PI Duan RD, Florence KA, Soppet DR;  
 PI WPI; 2000-679828/66.  
 DR Isolated nucleic acid molecule encoding a human secreted protein is  
 CC used in preventing, treating or ameliorating a medical condition -  
 CC Disclosure; Page 1057; 1065pp; English.  
 XX The polynucleotide sequences given in AAC99818 to AAC99977 encode the  
 CC human secreted proteins given in AAB56077 to AAB56362. Human secreted  
 CC proteins have activities based on the tissues and cells the genes are  
 CC expressed in. Examples of activities include: immunosuppressive;  
 CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;  
 CC vasotropic; cerebroprotective; neutrotropic; neuroprotective; antibacterial;  
 CC virucide; fungicide; and ophthalmological. The human secreted  
 CC polynucleotides and proteins can be used to prevent, treat or ameliorate  
 CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,  
 CC dogs, chickens or sheep. They are also used in diagnosing a pathological  
 CC condition or susceptibility to a pathological condition. Disorders which  
 CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid  
 CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or  
 CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular  
 CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders  
 CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and  
 CC fungi and ocular disorders e.g. corneal infection. The proteins can also  
 CC be used to aid wound healing and epithelial cell proliferation, to  
 CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplanted, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The proteins can also be used as a  
 CC food additive or preservative to increase or decrease storage  
 CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used  
 CC in the exemplification of the present invention.  
 XX Sequence 222 AA;  
 SQ  
 Query Match 59.4%; Score 38; DB 21; Length 222;  
 Best Local Similarity 53.8%; Pred. NO. 23;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 CLVLGSGARQLTF 13  
 ||| ||| :|  
 Db 105 cllcgssrttsf 117  
 RESULT 6  
 AAB56352  
 ID AAB56352 standard; Protein; 537 AA.  
 XX AAB56352;  
 XX 18-FEB-2002 (first entry)  
 DT Novel human diagnostic protein #26109.  
 DE

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 OS WO200175067-A2.  
 PN 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2000; 2000US-0540217.  
 XX 23-AUG-2000; 2000US-0649167.  
 PR (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 PI WPI; 2001-639362/73.  
 XX N-PSDB; AAS90305.  
 DR New isolated polynucleotide and encoded polypeptides, useful in  
 XX diagnostics, forensics, gene mapping, identification of mutations  
 XX responsible for genetic disorders or other traits and to assess  
 XX biodiversity -  
 XX Claim 20; SEQ ID No 56477; 103pp; English.  
 PS The invention relates to isolated polynucleotide (I) and  
 XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 537 AA;  
 SQ

Query Match 59.4%; Score 38; DB 22; Length 537;  
 Best Local Similarity 72.7%; Pred. No. 63;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CLVLGSGARQL 11  
 III I IIII:  
 Db 55 clvystsaqrql 65

RESULT 7  
 ABG08826  
 ID ABG08826 standard; Protein; 663 AA.  
 XX AC ABG08826;  
 XX 13-FEB-2002 (first entry)  
 DT Novel human diagnostic protein #8817.  
 XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 OS WO200175067-A2.  
 PN 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2000; 2000US-0540217.  
 XX 23-AUG-2000; 2000US-0649167.  
 PR (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 PI WPI; 2001-639362/73.  
 XX N-PSDB; AAS73013.  
 DR New isolated polynucleotide and encoded polypeptides, useful in  
 XX diagnostics, forensics, gene mapping, identification of mutations  
 XX responsible for genetic disorders or other traits and to assess  
 XX biodiversity -  
 XX Claim 20; SEQ ID No 39185; 103pp; English.  
 PS The invention relates to isolated polynucleotide (I) and  
 XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 663 AA;  
 SQ

Query Match 59.4%; Score 38; DB 22; Length 663;  
 Best Local Similarity 72.7%; Pred. No. 80;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CLVLGSGARQL 11  
 III I IIII:  
 Db 55 clvystsaqrql 65

RESULT 8  
 ABG29486  
 ID ABG29486 standard; Protein; 663 AA.  
 XX AC ABG29486;  
 XX 18-FEB-2002 (first entry)  
 DT Novel human diagnostic protein #29477.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW

KW food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 XX WO200175067-A2.  
 PN 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS93673.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX Claim 20; SEQ ID No 59845; 103pp; English.  
 PS The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 663 AA;

Query Match 59.4%; Score 38; DB 22; Length 663;  
 Best Local Similarity 72.7%; Pred. No. 80;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 CLVLGSGARQL 11  
 ||| | ||||  
 Db 55 clvystsrqi 65

RESULT 9  
 ABB32436  
 ID ABB32436 standard; Peptide; 34 AA.  
 XX  
 AC ABB32436;  
 XX  
 DT 01-FEB-2002 (first entry)  
 XX  
 DE Peptide #5087 encoded by breast cell single exon nucleic acid probe.  
 XX Human; microarray; single exon probe; gene expression; breast;  
 KW disease; cancer.

XX Homo sapiens.  
 OS WO200157271-A2.  
 PN 09-AUG-2001.  
 PD 30-JAN-2001; 2001WO-US00662.  
 PF 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-496933/54.  
 DR New spatially-addressable set of single exon nucleic acid probes,  
 PT useful for measuring gene expression in sample derived from human  
 PT breast, comprises number of single exon nucleic acid probes.  
 XX Claim 27; SEQ ID NO 15404; 327pp + sequence listing; English.  
 PS The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human breast and BT 474 cells. The method involves contacting  
 CC the probes with a collection of detectably labelled nucleic acids  
 CC derived from mRNA of human breast, and then measuring the label  
 CC bound to each probe of the microarray. The probes are useful for  
 CC verifying the expression of regions of genomic DNA predicted to  
 CC encode proteins. They are useful for gene discovery, and for  
 CC determining predisposition and/or prognosing breast disease. Gene  
 CC expression analysis is useful for assessing the toxicity of chemical  
 CC agents on cells. The microarray of this invention presents a far greater  
 CC diversity of probes for measuring gene expression, with far less bias  
 CC than expressed sequence tag microarrays. The method is suitable for  
 CC rapid production of functional information from genomic sequence. The  
 CC present sequence is a peptide encoded by a single exon nucleic acid  
 CC probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 34 AA;

Query Match 56.2%; Score 36; DB 22; Length 34;  
 Best Local Similarity 58.3%; Pred. No. 64;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 CLVLGSGARQLT 12  
 |||| | | |  
 Db 2 clvleggkrdls 13

RESULT 10  
 AAM58353  
 ID AAM58353 standard; Protein; 34 AA.  
 XX  
 AC AAM58353;  
 XX  
 DT 05-NOV-2001 (first entry)  
 XX  
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 30458.  
 XX Human; brain expressed exon; gene expression analysis; probe;  
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

KW epilepsy; cancer.

XX Homo sapiens.

XX WO200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX brains -

XX Example 4; SEQ ID NO: 30458; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX brain. They can be used to measure gene expression in brain cell samples,

XX which may enable the diagnosis and improved treatment of nervous system

XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

XX epilepsy and cancers. The present sequence is a protein encoded by one of

XX the probes of the invention.

XX Sequence 34 AA;

XX Query Match 56.2%; Score 36; DB 22; Length 34;

XX Best Local Similarity 58.3%; Pred. No. 6.4;

XX Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

XX 1 CLVLGSGARQLT 12

XX ||||| | | | |

XX 2 clvleggkrdls 13

XX RESULT 11

XX AAM18673

XX ID AAM18673 standard; Protein; 34 AA.

XX AAM18673;

XX 12-OCT-2001 (first entry)

XX Peptide #5107 encoded by probe for measuring cervical gene expression.

XX Probe; human; microarray; gene expression; cervical epithelial cell;

XX cervical cancer.

XX Homo sapiens.

XX WO200157278-A2.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human cervical epithelial cells -

XX Claim 27; SEQ ID No 23499; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes

XX (SENP; see AAI10068-AAI28459). The present sequence is a peptide encoded

XX by one such probe. The SENPs are derived from human Hela cells. The SENPs

XX can be used to produce a single exon microarray, which can be used for

XX measuring human gene expression in a sample derived from human cervical

XX epithelial cells. By measuring gene expression, the probes are therefore

XX useful in grading and/or staging of diseases of the cervix, notably

XX cervical cancer.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 34 AA;

XX Query Match 56.2%; Score 36; DB 22; Length 34;

XX Best Local Similarity 58.3%; Pred. No. 6.4;

XX Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

XX 1 CLVLGSGARQLT 12

XX ||||| | | | |

XX 2 clvleggkrdls 13

XX RESULT 12

XX AAU39475

XX ID AAU39475 standard; Protein; 147 AA.

XX AC AAU39475;

XX 13-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #371.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;

XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;

XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;

XX dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

DR N-PSDB; AAS59507.  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 XX Example 1; SEQ ID No 670; 1059pp; English.  
 PS  
 XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 147 AA;

Query Match 54.7%; Score 35; DB 22; Length 147;  
 Best Local Similarity 63.6%; Pred. No. 55;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVLGSRQLT 12  
 DB 107 lvtgtahqlt 117  
 ||:|:| |

RESULT 13  
 AAG84910  
 ID AAG84910 standard; Protein; 284 AA.  
 XX  
 AC AAG84910;  
 XX  
 XX 11-SEP-2001 (first entry)  
 XX  
 DE Shrimp white spot Bacilliform virus (WSBV) protein 1.  
 XX  
 KW Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;  
 KW antiviral agent; gene expression; antisense construct;  
 KW transgenic viral resistant shrimp.  
 XX  
 OS White spot syndrome virus.  
 XX  
 XX WO200138351-A2.  
 XX  
 XX 31-MAY-2001.  
 XX  
 XX 08-NOV-2000; 2000WO-US28888.  
 PF  
 XX 24-NOV-1999; 99CN-0124717.  
 PR  
 XX (PENY-) PE CORP NY.  
 PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.  
 PA (SINO-) SINOGENOMAX CO LTD.  
 XX  
 XX Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;  
 PI  
 XX WPI; 2001-355877/37.  
 DR  
 DR N-PSDB; AAG84910.

XX Primary nucleotide sequence of the shrimp white spot Bacilliform virus  
 PT (WSBV), useful for producing viral polypeptides that can be used to  
 PT screen for agents that are useful for treating WSBV infection -  
 XX  
 XX Claim 1; Figure 3; 626pp; English.  
 PS  
 XX The invention provides the primary nucleotide sequence of the WSBV genome  
 CC (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and  
 CC encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences  
 CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid  
 CC molecules and proteins of the invention are useful for diagnosis and  
 CC monitoring viral infection, in screens for antiviral agents and for  
 CC monitoring viral gene expression or activity during a treatment regimen.  
 CC The nucleic acid molecules are also useful as antisense constructs to  
 CC control viral gene expression in infected cells and tissues and to create  
 CC transgenic viral resistant shrimp.  
 XX  
 XX Sequence 284 AA;  
 SQ  
 Query Match 54.7%; Score 35; DB 22; Length 284;  
 Best Local Similarity 60.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 4 LSGSRQLTF 13  
 DB 199 lsgksrqlty 208  
 ||:|:| |

RESULT 14  
 AAG91501  
 ID AAG91501 standard; Protein; 334 AA.  
 XX  
 AC AAG91501;  
 XX  
 XX 26-SEP-2001 (first entry)  
 DT  
 XX  
 XX C glutamicum protein fragment SEQ ID NO: 5255.  
 DE  
 DE Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 KW  
 XX Corynebacterium glutamicum.  
 OS  
 XX EP1108790-A2.  
 XX  
 XX 20-JUN-2001.  
 PD  
 XX  
 XX 18-DEC-2000; 2000EP-0127688.  
 PF  
 XX  
 XX 16-DEC-1999; 99JP-0377484.  
 PR  
 XX 07-APR-2000; 2000JP-0159162.  
 PR  
 XX 03-AUG-2000; 2000JP-0280988.  
 PR  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA  
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 PI  
 XX WPI; 2001-376931/40.  
 DR  
 DR N-PSDB; AAG91501.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 XX Claim 17; SEQ ID NO: 5255; 246pp + Sequence Listing: English.  
 PS  
 XX The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a

CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Corynebacterium, and identifying a homologue of a gene derived  
 CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX  
 XX Sequence 334 AA;

Query Match 54.7%; Score 35; DB 22; Length 334;  
 Best Local Similarity 63.6%; Pred. NO. 1.4e+02;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 VLSGSARQLTF 13  
 I::||| |||  
 Db 237 vitgsatdlf 247

RESULT 15  
 AAB79276  
 ID AAB79276 standard; Protein; 334 AA.

AC AAB79276;  
 DT 30-APR-2001 (first entry)

DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:68.

XX Corynebacterium glutamicum; carbon metabolism and energy production;  
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;  
 KW fine chemical production; organic acid; proteinogenic amino acid;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KW carboxylate; aromatic compound; vitamin; cofactor; polyketide; enzyme;  
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study.

XX Corynebacterium glutamicum.

OS  
 XX WO200100844-A2.

PN  
 XX 04-JAN-2001.

PD  
 XX 23-JUN-2000; 2000WO-IB00943.

PF  
 XX 25-JUN-1999; 99US-0141031.

PR 08-JUL-1999; 99DE-1031412.

PR 08-JUL-1999; 99DE-1031413.

PR 08-JUL-1999; 99DE-1031419.

PR 08-JUL-1999; 99DE-1031420.

PR 08-JUL-1999; 99DE-1031424.

PR 08-JUL-1999; 99DE-1031428.

PR 08-JUL-1999; 99DE-1031431.

PR 08-JUL-1999; 99DE-1031433.

PR 08-JUL-1999; 99DE-1031434.

PR 08-JUL-1999; 99DE-1031510.

PR 08-JUL-1999; 99DE-1031562.

PR 08-JUL-1999; 99DE-1031634.

PR 09-JUL-1999; 99DE-1032180.

PR 09-JUL-1999; 99DE-1032227.

PR 09-JUL-1999; 99DE-1032230.

PR 09-JUL-1999; 99US-0143208.

PR 03-SEP-1999; 99DE-1042086.  
 PR 03-SEP-1999; 99DE-1042087.  
 PR 03-SEP-1999; 99DE-1042088.  
 PR 03-SEP-1999; 99DE-1042095.  
 PR 03-SEP-1999; 99DE-1042123.  
 PR 03-SEP-1999; 99DE-1042125.

XX (BADI ) BASF AG.

PA Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

PI WPI; 2001-061975/07.

XX N-PSDB; AAF71393.

DR New isolated Corynebacterium glutamicum nucleic acid encoding a sugar  
 XX metabolism and oxidative phosphorylation protein for production or  
 PT modulation of production of fine chemicals e.g. amino acids,  
 PT carbohydrates or enzymes -

XX Claim 20; Page 238; 1246pp; English.

XX AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar  
 CC metabolism and oxidative phosphorylation (SMP) proteins given in  
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and  
 CC energy production. The C. glutamicum SMP gene can be used in vectors  
 CC (II) for expression in host cells and production or modulation of  
 CC production of fine chemicals, such as, an organic acid, a proteinogenic  
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,  
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty  
 CC acid, a diol, a carboxylate, an aromatic compound, a vitamin, a  
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins  
 CC (III) encoded by them are used for diagnosing the presence or activity of  
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells  
 CC containing them are used to map genomes of organisms related to  
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,  
 CC in evolutionary studies, in determining SMP protein regions required  
 CC for function, in modulating SMP protein activity, in modulating the  
 CC metabolism of sugars, and in modulating high-energy molecule production  
 CC in a cell (i.e. ATP, NADPH).

XX Sequence 334 AA;

Query Match 54.7%; Score 35; DB 22; Length 334;

Best Local Similarity 63.6%; Pred. NO. 1.4e+02;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 VLSGSARQLTF 13

Db 237 vitgsatdlf 247

Search completed: June 6, 2002, 06:17:20

Job time: 124 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:15:16 ; Search time 21.84 Seconds  
(without alignments)  
14.539 Million cell updates/sec

Title: US-08-881-509-6

Perfect score: 64

Sequence: 1 CLVLSGARQLTF 13

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	53.1	28	3	US-08-963-121C-11
2	34	53.1	28	4	US-09-543-513-11
3	34	53.1	28	5	PCT-US95-04803-12
4	34	53.1	59	4	US-08-905-223-305
5	34	53.1	106	6	5204252-6
6	34	53.1	269	3	US-08-926-842B-59
7	34	53.1	297	3	US-09-173-361-1
8	34	53.1	297	4	US-09-420-915-1
9	33.5	52.3	115	2	US-08-485-449-4
10	33.5	52.3	376	2	US-08-485-449-5
11	33.5	52.3	389	2	US-08-485-449-2
12	33.5	52.3	389	2	US-08-485-449-6
13	33.5	52.3	389	2	US-08-485-449-7
14	33	51.6	376	3	US-09-025-691-3
15	33	51.6	762	4	US-09-228-986-114
16	32.5	50.8	1500	4	US-09-323-472A-2
17	32.5	50.8	1500	4	US-09-323-472A-4
18	32.5	50.8	1500	4	US-09-323-472A-12
19	32.5	50.8	1500	4	US-09-323-472A-14
20	32	50.0	252	4	US-09-199-637A-176
21	31	48.4	140	1	US-08-664-596B-2
22	31	48.4	140	2	US-08-738-367-2
23	31	48.4	165	3	US-08-685-808-4
24	31	48.4	165	4	US-08-505-860C-4
25	31	48.4	202	1	US-08-469-486-56
26	31	48.4	202	2	US-08-469-658-56
27	31	48.4	240	2	US-08-738-367-7

28	31	48.4	255	1	US-08-459-264-4	Sequence 4, Appl
29	31	48.4	255	1	US-08-459-263-4	Sequence 4, Appl
30	31	48.4	807	4	US-08-855-910-9	Sequence 9, Appl
31	31	48.4	824	3	US-08-221-750A-3	Sequence 3, Appl
32	31	48.4	844	3	US-09-029-267-20	Sequence 20, Appl
33	31	48.4	890	1	US-08-472-934-8	Sequence 8, Appl
34	31	48.4	890	2	US-08-323-460A-8	Sequence 8, Appl
35	31	48.4	890	2	US-08-461-146C-8	Sequence 8, Appl
36	31	48.4	890	3	US-08-461-145C-8	Sequence 8, Appl
37	31	48.4	890	4	US-08-628-829-12	Sequence 12, Appl
38	31	48.4	1247	1	US-08-472-934-10	Sequence 10, Appl
39	31	48.4	1247	2	US-08-323-460A-10	Sequence 10, Appl
40	31	48.4	1247	2	US-08-461-146C-10	Sequence 10, Appl
41	31	48.4	1247	3	US-08-461-145C-10	Sequence 10, Appl
42	31	48.4	1597	4	US-09-423-890-13	Sequence 13, Appl
43	31	48.4	1597	4	US-08-628-829-14	Sequence 14, Appl
44	30	46.9	107	1	US-08-276-852-107	Sequence 107, App
45	30	46.9	107	1	US-08-899-575-107	Sequence 107, App

#### ALIGNMENTS

RESULT 1  
US-08-963-121C-11  
; Sequence 11, Application US/08963121C  
; Patent No. 6084087  
; GENERAL INFORMATION:  
; APPLICANT: Friedman, Steven M  
; APPLICANT: Crow, Mary K  
; APPLICANT: Yi, Y.  
; APPLICANT: Tumang, Joseph  
; APPLICANT: Sun, Guang-Rong  
; TITLE OF INVENTION: Conserved T-Cell Receptor Sequences  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: US  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/963,121C  
; FILING DATE: October 28, 1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/427,009  
; FILING DATE: April, 24, 1995  
; APPLICATION NUMBER: 08/229,285  
; FILING DATE: April 18, 1994  
; APPLICATION NUMBER: 07/766,751, Patent No. 6084087 5,480,895  
; FILING DATE: September 27, 1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, S. Peter  
; REGISTRATION NUMBER: 25,351  
; REFERENCE/DOCKET NUMBER: 5983/17499-US1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide

FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: Va3.1-aa  
US-08-963-121C-11

Query Match 53.1%; Score 34; DB 3; Length 28;  
Best Local Similarity 53.8%; Pred. No. 3.1; Mismatches 5; Indels 0; Gaps 0;

Matches 7; Conservative 1; Mismatches 0; Gaps 0;

QY 1 CLVLGSGARQLTF 13  
| | | | :|||  
Db 1 CATLGGSNYKLTF 13

RESULT 2  
US-09-543-513-11  
; Sequence 11, Application US/09543513  
; Patent No. 6303750  
; GENERAL INFORMATION:  
; APPLICANT: Friedman, Steven M  
; APPLICANT: Crow, Mary K  
; APPLICANT: Yi, Y.  
; APPLICANT: Tumang, Joseph  
; APPLICANT: Sun, Guang-Rong  
; TITLE OF INVENTION: Conserved T-Cell Receptor Sequences  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: US  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/543,513  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/963,121  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, S. Peter  
; REGISTRATION NUMBER: 25,351  
; REFERENCE/DOCKET NUMBER: 5983/17499-US1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: Va3.1-aa  
US-09-543-513-11

Query Match 53.1%; Score 34; DB 4; Length 28;  
Best Local Similarity 53.8%; Pred. No. 3.1;

Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 CLVLGSGARQLTF 13  
| | | | :|||  
Db 1 CATLGGSNYKLTF 13

RESULT 3  
PCT-US95-04803-12  
; Sequence 12, Application PC/TUS9504803  
; GENERAL INFORMATION:  
; APPLICANT: New York Society For the ruptured and  
; APPLICANT: Crippled Maintaining The Hospital for  
; APPLICANT: Special Surgery  
; APPLICANT: INVENTORS: Friedman, Steven M  
; APPLICANT: Crow, Mary K  
; APPLICANT: Yi, Y.  
; APPLICANT: Tumang, Joseph  
; APPLICANT: Sun, Guang-Rong  
; TITLE OF INVENTION: Conserved T-Cell Receptor Sequences  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: US  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/04803  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, S. Peter  
; REGISTRATION NUMBER: 25,351  
; REFERENCE/DOCKET NUMBER: 5983/09449  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: Va3.1-aa  
PCT-US95-04803-12

Query Match 53.1%; Score 34; DB 5; Length 28;  
Best Local Similarity 53.8%; Pred. No. 3.1;  
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 CLVLGSGARQLTF 13  
| | | | :|||  
Db 1 CATLGGSNYKLTF 13

RESULT 4  
US-08-905-223-305  
; Sequence 305, Application US/08905223  
; Patent No. 622029



```

; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste D.
; APPLICANT: Duclert, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
; NUMBER OF SEQUENCES: 503
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,223
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 305:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 59 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Brain
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: -21--1
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 6.9
; OTHER INFORMATION: seq LLAGCSLLPLGLWQ/HL
; US-08-905-223-305

```

```

Query Match 53.1%; Score 34; DB 4; Length 59;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 1 CLVLGSGARQLT 12
| | | | | | | |
Db 36 CTVCSSSEQIS 47

```

```

RESULT 5
5204252-6
; Patent No. 5204252
; APPLICANT: CRREGG, JAMES M.; CLEESON, MARTIN A.; HAAS, LISA
; PICATAGGIO, STEPHEN
; TITLE OF INVENTION: CANDIDA TROPICALIS TRANSFORMATION SYSTEM
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/386,837
; FILING DATE: 27-JUL-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 308,481
; FILING DATE: 08-FEB-1989
; SEQ ID NO: 6
; LENGTH: 106
5204252-6

```

```

Query Match 53.1%; Score 34; DB 6; Length 106;
Best Local Similarity 41.7%; Pred. No. 14;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 1 CLVLGSGARQLT 12
| | | | | | | |
Db 20 CIVMAGGAQAVT 31

```

```

RESULT 6
US-08-926-842B-59
; Sequence 59, Application US/08926842B
; Patent No. 6030807
; GENERAL INFORMATION:
; APPLICANT: Sa-No. 6030807ueira, Isabel
; APPLICANT: de Lencastre, Herminia
; TITLE OF INVENTION: HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,842B
; FILING DATE: 10-SEP-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-089 N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Bacillus subtilis
; FEATURE:
; OTHER INFORMATION: /product="araL"
US-08-926-842B-59

```

```

Query Match 53.1%; Score 34; DB 3; Length 269;
Best Local Similarity 77.8%; Pred. No. 41;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 2 LVLGSGARQ 10
| | | | | | | |
Db 233 LVLGSGARQ 241

```

```

RESULT 7
US-09-173-581-1
; Sequence 1, Application US/09173581A
; Patent No. 6013455
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Tang, Y. Tom

```

; Sequence 4, Application US/08485449  
; Patent No. 5824789  
; GENERAL INFORMATION:  
; APPLICANT: VANDENBERG, DAVID  
; TITLE OF INVENTION: HEMATOPOIETIC GROWTH FACTORS, NUCLEOTIDE  
; TITLE OF INVENTION: SEQUENCE ENCODING GROWTH FACTORS AND METHODS OF USE  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,449  
; FILING DATE:  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KONSKI, ANTOINETTE F.  
; REGISTRATION NUMBER: 34,202  
; REFERENCE/DOCKET NUMBER: 20296-20035.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 813-5600  
; TELEFAX: (415) 494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 115 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-485-449-4

Query Match 52.3%; Score 33.5; DB 2; Length 115;  
Best Local Similarity 75.0%; Pred. NO. 20;  
Matches 9; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 CLVLGSG-SARQL 11  
|| ||| | |||  
Db 48 CLTSLGSLSKRQL 59

RESULT 10  
US-08-485-449-5  
; Sequence 5, Application US/08485449  
; Patent No. 5824789  
; GENERAL INFORMATION:  
; APPLICANT: VANDENBERG, DAVID  
; TITLE OF INVENTION: HEMATOPOIETIC GROWTH FACTORS, NUCLEOTIDE  
; TITLE OF INVENTION: SEQUENCE ENCODING GROWTH FACTORS AND METHODS OF USE  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Yue, Henry  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Gorgone, Gina  
; APPLICANT: Azimzai, Yalda  
; APPLICANT: Lu, Aina  
; TITLE OF INVENTION: Protein Kinase Homologs  
; FILE REFERENCE: PF-0614 US  
; CURRENT APPLICATION NUMBER: US/09/173,581A  
; CURRENT FILING DATE: 1998-10-15  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PERL Program  
; SEQ ID NO 1  
; LENGTH: 297  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE: -  
; OTHER INFORMATION: 119819  
; US-09-173-581-1

Query Match 53.1%; Score 34; DB 3; Length 297;  
Best Local Similarity 60.0%; Pred. No. 46;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CLVLGSGARQ 10  
| : |||| : |  
Db 93 CSILGSSNQ 102

RESULT 8  
US-09-420-915-1  
; Sequence 1, Application US/09420915  
; Patent No. 6284947  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Yue, Henry  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Gorgone, Gina  
; APPLICANT: Azimzai, Yalda  
; APPLICANT: Lu, Aina  
; TITLE OF INVENTION: Protein Kinase Homologs  
; FILE REFERENCE: PF-0614 US  
; CURRENT APPLICATION NUMBER: US/09/420,915  
; CURRENT FILING DATE: 1999-10-20  
; EARLIER FILING DATE: 1998-10-15  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PERL Program  
; SEQ ID NO 1  
; LENGTH: 297  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE: -  
; OTHER INFORMATION: 119819  
; US-09-420-915-1

Query Match 53.1%; Score 34; DB 4; Length 297;  
Best Local Similarity 60.0%; Pred. No. 46;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CLVLGSGARQ 10  
| : |||| : |  
Db 93 CSILGSSNQ 102

RESULT 9  
US-08-485-449-4

APPLICATION NUMBER: US/08/485,449  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: KONSKI, ANTOINETTE F.  
REGISTRATION NUMBER: 34,202  
REFERENCE/DOCKET NUMBER: 20296-20035.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 376 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-485-449-5

Query Match 52.3%; Score 33.5; DB 2; Length 376;  
Best Local Similarity 75.0%; Pred. No. 76;  
Matches 9; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 CLVLGS-SARQL 11  
||| ||| |||  
DB 44 CLTSLGSLSKROL 55

RESULT 11  
US-08-485-449-2  
Sequence 2, Application US/08485449  
Patent No. 5824789  
GENERAL INFORMATION:  
APPLICANT: VANDENBERG, DAVID  
TITLE OF INVENTION: HEMATOPOIETIC GROWTH FACTORS, NUCLEOTIDE  
TITLE OF INVENTION: SEQUENCE ENCODING GROWTH FACTORS AND METHODS OF USE  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,449  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: KONSKI, ANTOINETTE F.  
REGISTRATION NUMBER: 34,202  
REFERENCE/DOCKET NUMBER: 20296-20035.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 389 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-485-449-2

Query Match 52.3%; Score 33.5; DB 2; Length 389;

Best Local Similarity 75.0%; Pred. No. 79;  
Matches 9; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 CLVLGS-SARQL 11  
||| ||| |||  
DB 48 CLTSLGSLSKROL 59

RESULT 12  
US-08-485-449-6  
Sequence 6, Application US/08485449  
Patent No. 5824789  
GENERAL INFORMATION:  
APPLICANT: VANDENBERG, DAVID  
TITLE OF INVENTION: HEMATOPOIETIC GROWTH FACTORS, NUCLEOTIDE  
TITLE OF INVENTION: SEQUENCE ENCODING GROWTH FACTORS AND METHODS OF USE  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,449  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: KONSKI, ANTOINETTE F.  
REGISTRATION NUMBER: 34,202  
REFERENCE/DOCKET NUMBER: 20296-20035.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 389 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-485-449-6

Query Match 52.3%; Score 33.5; DB 2; Length 389;  
Best Local Similarity 75.0%; Pred. No. 79;  
Matches 9; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 CLVLGS-SARQL 11  
||| ||| |||  
DB 48 CLTSLGSLSKROL 59

RESULT 13  
US-08-485-449-7  
Sequence 7, Application US/08485449  
Patent No. 5824789  
GENERAL INFORMATION:  
APPLICANT: VANDENBERG, DAVID  
TITLE OF INVENTION: HEMATOPOIETIC GROWTH FACTORS, NUCLEOTIDE  
TITLE OF INVENTION: SEQUENCE ENCODING GROWTH FACTORS AND METHODS OF USE  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road

;; CITY: Palo Alto  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94304-1018  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/485,449  
;; FILING DATE:  
;; CLASSIFICATION: 536  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: KONSIT, ANTOINETTE F.  
;; REGISTRATION NUMBER: 34,202  
;; REFERENCE/DOCKET NUMBER: 20296-20035.00  
;; TELEPHONE: (415) 813-5600  
;; TELEFAX: (415) 494-0792  
;; TELEX: 706141  
;; INFORMATION FOR SEQ ID NO: 7:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 389 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; US-08-485-449-7

Query Match 52.3%; Score 33.5; DB 2; Length 389;  
Best Local Similarity 75.0%; Pred. No. 79;  
Matches 9; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 CLVLGSG-SARQL 11  
Db 48 CLTGLSLKRQL 59

RESULT 14  
US-09-025-691-3  
; Sequence 3, Application US/09025691  
; Patent No. 6069299  
; GENERAL INFORMATION:  
; APPLICANT: Broadway, Roxanne M.  
; APPLICANT: Harman, Gary E.  
; TITLE OF INVENTION: FUNGUS AND INSECT CONTROL WITH  
; TITLE OF INVENTION: CHITINOLYTIC ENZYMES  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/025,691  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/20120  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (716) 263-1304  
; TELEFAX: (716) 263-1600  
; INFORMATION FOR SEQ ID NO: 3:

;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 376 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-09-025-691-3

Query Match 51.6%; Score 33; DB 3; Length 376;  
Best Local Similarity 63.6%; Pred. No. 95;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 LVLGSGSARQLT 12  
Db 30 LVTSGSAEKIT 40

RESULT 15  
US-09-228-986-114  
; Sequence 114, Application US/09228986  
; Patent No. 6359198  
; GENERAL INFORMATION:  
; APPLICANT: Strabala, Timothy  
; APPLICANT: Nieuwenhuizen, Niels  
; TITLE OF INVENTION: Compositions Isolated from Plant Cells  
; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling  
; FILE REFERENCE: 11000/1020  
; CURRENT APPLICATION NUMBER: US/09/228,986  
; CURRENT FILING DATE: 1999-01-12  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 114  
; LENGTH: 762  
; TYPE: PRT  
; ORGANISM: Pinus radiata  
; US-09-228-986-114

Query Match 51.6%; Score 33; DB 4; Length 762;  
Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 LVLGSGSARQLTF 13  
Db 673 LLQSGSAFQLVF 684

Search completed: June 6, 2002, 06:15:47  
Job time: 31 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:21:22 ; Search time 14.13 Seconds  
(without alignments)  
88.405 Million cell updates/sec

Title: US-08-881-509-6

Perfect score: 64

Sequence: 1 CIVLSSGSARQLTF 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 1824

Minimum DB seq length: 0  
Maximum DB seq length: 13

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR71.\*

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26.5	41.4	13	2 S57571	T cell receptor al
2	26	40.6	13	2 S47361	T-cell antigen rec
3	22	34.4	10	2 S23771	T-cell receptor al
4	22	34.4	13	2 PH1772	T cell receptor al
5	21	32.8	10	2 PT0212	T-cell receptor al
6	20	31.2	8	2 S08996	hypertrehalosemic
7	20	31.2	8	2 B49823	adipokinetic hormo
8	20	31.2	8	2 B44960	neuropeptide led-C
9	20	31.2	8	2 A33995	adipokinetic hormo
10	20	31.2	8	2 S11545	adipokinetic hormo
11	20	31.2	9	2 A24244	hypertrehalosemic
12	20	31.2	10	2 JC1416	hypertrehalosemic/
13	20	31.2	10	2 S09138	hypertrehalosemic
14	20	31.2	10	2 A31571	hypertrehalosemic
15	20	31.2	10	2 B33995	hypertrehalosemic
16	20	31.2	13	2 S32473	hypertrehalosemic
17	20	31.2	13	2 B56864	hypertrehalosemic
18	19	29.7	6	2 PT0368	lymphadamide 3 - g
19	19	29.7	10	2 S48182	dipeptidyl-peptida
20	19	29.7	10	2 A47364	T-cell receptor be
21	19	29.7	11	2 S23364	bacterioferritin -
22	19	29.7	11	2 S61797	placental lactogen
23	19	29.7	12	2 A37933	T-cell receptor al
24	18	28.1	6	2 PT0280	T-cell-specific tr
25	18	28.1	10	2 G60787	Ig lambda chain CRD
26	18	28.1	10	2 E60787	sperm-activating p
27	18	28.1	10	2 E60589	sperm-activating p
28	18	28.1	10	2 S70251	nitrogenase (EC 1.
29	18	28.1	10	2 C41946	T-cell receptor ga

## ALIGNMENTS

## RESULT 1

S57371  
T cell receptor alpha chain V-J region (clone pp7 and others) - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 19-Oct-1995 #sequence\_revision 17-Nov-1995 #text\_change 05-Nov-1999  
C:Accession: S57571; S57573; S57576  
R:Burrows, S.R.; Silins, S.L.; Moss, D.J.; Khanna, R.; Misko, I.S.; Argast, V.P.  
submitted to the EMBL Data Library, June 1995  
A:Description: T cell receptor repertoire for a viral epitope in humans is diversifie  
A:Reference number: S57494  
A:Accession: S57571  
A:Molecule type: mRNA  
A:Residues: 1-13 <BUR>  
A:Cross-references: EMBL:249948; NID:9887496; PIDN:CAA90221.1; PID:9887497  
A:Experimental source: clone pp7  
A:Accession: S57573  
A:Molecule type: mRNA  
A:Residues: 1-13 <BUL>  
A:Cross-references: EMBL:249950; NID:9887500; PIDN:CAA90221.1; PID:9887501  
A:Experimental source: clone TF1  
A:Accession: S57576  
A:Molecule type: mRNA  
A:Residues: 1-13 <BUW>  
A:Cross-references: EMBL:249952; NID:9887512; PIDN:CAA90223.1; PID:9887513  
A:Experimental source: clone RL16  
C:Keywords: T-cell receptor

Query Match 41.4%; Score 26.5; DB 2; Length 13;  
Best Local Similarity 53.8%; Pred. No. 1.1e+02;  
Matches 7; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 CIVLSSGSARQLTF 13  
DB 1 CAVLFGN-EKLTFF 12

## RESULT 2

S47361  
T-cell antigen receptor VJ junction beta chain - human  
C:Species: Homo sapiens (man)  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 05-Nov-1999  
C:Accession: S47361  
R:Lehner, P.J.

submitted to the EMBL Data Library, August 1994

A:Description: Human HLA-A0201 restricted recognition of influenza A is dominated by  
A:Reference number: S47355  
A:Accession: S47361  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-13 <LEH>  
A:Cross-references: EMBL:235685; NID:9527459; PIDN:CAA84754.1; PID:9527460

C;Keywords: T-cell receptor

Query Match 40.6%; Score 26; DB 2; Length 13;  
Best Local Similarity 46.2%; Pred. No. 1.3e+02;  
Matches 6; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 CLVLGSGARQLTF 13  
| | | | |  
Db 1 CSVLQSGPVEQVF 13

RESULT 3

S23371  
T-cell receptor alpha chain J region - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 22-Nov-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C;Accession: S23371  
R;Pluschke, G.; Ricken, G.; Taube, H.; Kroninger, S.; Melchers, I.; Peter, H.H.; Eichman  
Eur. J. Immunol. 21, 2749-2754, 1991  
A;Title: Biased T cell receptor V(alpha) region repertoire in the synovial fluid of rheu  
A;Reference number: S23364; MUID:92037820  
A;Accession: S23371  
A;Status: preliminary; translation not shown  
A;Molecule type: mRNA  
A;Residues: 1-10 <PLU>  
A;Cross-references: EMBL:X58166  
C;Keywords: T-cell receptor

Query Match 34.4%; Score 22; DB 2; Length 10;  
Best Local Similarity 55.6%; Pred. No. 6.3e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 SSGARQLTF 13  
| | | | |  
Db 2 SGEAGKSTF 10

RESULT 4

PH1772  
T cell receptor alpha chain V region (clone 2V alpha 23-3) - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999  
C;Accession: PH1772  
R;Porcellini, S.; Yockey, C.E.; Brenner, M.B.; Balk, S.P.  
J. Exp. Med. 178, 1-16, 1993  
A;Title: Analysis of T cell antigen receptor (TCR) expression by human peripheral blood  
A;Reference number: PH1754; MUID:93301585  
A;Accession: PH1772  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-13 <POR>

Query Match 34.4%; Score 22; DB 2; Length 13;  
Best Local Similarity 83.3%; Pred. No. 8.2e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 ARQLTF 13  
| | | | |  
Db 7 AGQLTF 12

RESULT 5

PT0212  
T-cell receptor alpha chain V-J region (4-1-E.2) - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 30-May-1997  
C;Accession: PT0212  
R;Nakano, N.; Kikutani, H.; Nishimoto, H.; Kishimoto, T.  
J. Exp. Med. 173, 1091-1097, 1991

A;Title: T cell receptor V gene usage of islet beta cell-reactive T cells is not rest  
A;Reference number: PT0209; MUID:91217621  
A;Accession: PT0212  
A;Molecule type: mRNA  
A;Residues: 1-10 <NAK>  
C;Keywords: T-cell receptor

Query Match 32.8%; Score 21; DB 2; Length 10;  
Best Local Similarity 44.4%; Pred. No. 9.9e+02;  
Matches 4; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 CLVLGSGAR 9  
| | | | |  
Db 1 CAVAGGADR 9

RESULT 6

S08996  
hypertrehalosemic hormone II - oriental cockroach  
N;Alternate names: Pea-CAH-II  
C;Species: Blattia orientalis (oriental cockroach)  
C;Date: 30-Jun-1992 #sequence\_revision 24-Oct-1997 #text\_change 31-Oct-1997  
C;Accession: S08996  
R;Gaede, G.; Rinehart, K.L.  
Biol. Chem. Hoppe-Seyler 371, 345-354, 1990  
A;Title: Primary structures of hypertrehalosemic neuropeptides isolated from the cor  
entails and of the stick insect Extatosoma tiaratum assigned by tandem fast atom bomb  
A;Reference number: S08995; MUID:90253659  
A;Accession: S08996  
A;Molecule type: protein  
A;Residues: 1-8 <GA>  
A;Note: the amino-terminal residue forms pyrrolidone carboxylic acid; therefore, we h  
C;Superfamily: adipokinetic hormone  
C;Keywords: amidated carboxyl end; hormone; neuropeptide; pyroglutamic acid  
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
F;8/Modified site: amidated carboxyl end (Trp) #status experimental

Query Match 31.2%; Score 20; DB 2; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 QLTF 13  
| | | | |  
Db 1 QLTF 4

RESULT 7

B49823  
adipokinetic hormone II - American cockroach  
N;Alternate names: neuropeptide M-II; periplanetin CC-1  
C;Species: Periplaneta americana (American cockroach)  
C;Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 07-May-1999  
C;Accession: B49823; A05170  
R;Scarborough, R.M.; Jamieson, G.C.; Kalish, F.; Kramer, S.J.; McEnroe, G.A.; Miller,  
Proc. Natl. Acad. Sci. U.S.A. 81, 5575-5579, 1984  
A;Title: Isolation and primary structure of two peptides with cardioacceleratory and  
A;Reference number: A49823; MUID:84298179  
A;Accession: B49823  
A;Molecule type: protein  
A;Residues: 1-8 <SCA>  
R;Witten, J.L.; Schafer, M.H.; O'Shea, M.; Cook, J.C.; Hemling, M.E.; Rinehart Jr.,  
Biochem. Biophys. Res. Commun. 124, 350-358, 1984  
A;Title: Structures of two cockroach neuropeptides assigned by fast atom bombardment  
A;Reference number: A90118; MUID:85046530  
A;Accession: A05170  
A;Molecule type: protein  
A;Residues: 'E', 2-8 <MIT>  
C;Superfamily: adipokinetic hormone  
C;Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyrogluta  
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
F;8/Modified site: amidated carboxyl end (Trp) #status experimental

C;Keywords: T-cell receptor

Query Match 40.6%; Score 26; DB 2; Length 13;  
Best Local Similarity 46.2%; Pred. No. 1.3e+02;  
Matches 6; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 CLVLGSGARQLTF 13  
| | | | |  
Db 1 CSVLQSGPVEQVF 13

RESULT 3

S23371  
T-cell receptor alpha chain J region - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 22-Nov-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C;Accession: S23371  
R;Pluschke, G.; Ricken, G.; Taube, H.; Kroninger, S.; Melchers, I.; Peter, H.H.; Eichman  
Eur. J. Immunol. 21, 2749-2754, 1991  
A;Title: Biased T cell receptor V(alpha) region repertoire in the synovial fluid of rheu  
A;Reference number: S23364; MUID:92037820  
A;Accession: S23371  
A;Status: preliminary; translation not shown  
A;Molecule type: mRNA  
A;Residues: 1-10 <PLU>  
A;Cross-references: EMBL:X58166  
C;Keywords: T-cell receptor

Query Match 34.4%; Score 22; DB 2; Length 10;  
Best Local Similarity 55.6%; Pred. No. 6.3e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 SSGARQLTF 13  
| | | | |  
Db 2 SGEAGKSTF 10

RESULT 4

PH1772  
T cell receptor alpha chain V region (clone 2V alpha 23-3) - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999  
C;Accession: PH1772  
R;Porcellini, S.; Yockey, C.E.; Brenner, M.B.; Balk, S.P.  
J. Exp. Med. 178, 1-16, 1993  
A;Title: Analysis of T cell antigen receptor (TCR) expression by human peripheral blood  
A;Reference number: PH1754; MUID:93301585  
A;Accession: PH1772  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-13 <POR>

Query Match 34.4%; Score 22; DB 2; Length 13;  
Best Local Similarity 83.3%; Pred. No. 8.2e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 ARQLTF 13  
| | | | |  
Db 7 AGQLTF 12

RESULT 5

PT0212  
T-cell receptor alpha chain V-J region (4-1-E.2) - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 30-May-1997  
C;Accession: PT0212  
R;Nakano, N.; Kikutani, H.; Nishimoto, H.; Kishimoto, T.  
J. Exp. Med. 173, 1091-1097, 1991

Query Match 31.2%; Score 20; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 QLTF 13  
 ||||  
 Db 1 QLTF 4

RESULT 8  
 B44960  
 neuropeptide Led-CC-II - Colorado potato beetle  
 C:Species: Leptinotarsa decemlineata (Colorado potato beetle)  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
 C:Accession: B44960  
 R:Gaede, G.; Kellner, R.  
 Reptides 10, 1287-1289, 1989  
 A:Title: The metabolic neuropeptides of the corpus cardiacum from the potato beetle and  
 A:Reference number: A44960; MUID:90160053  
 A:Accession: B44960  
 A:Molecule type: protein  
 A:Residues: 1-8 <GAE>  
 C:Superfamily: adipokinetic hormone  
 C:Keywords: blocked carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
 F:8/Modified site: blocked carboxyl end (Trp) (probably amidated) #status experimental

Query Match 31.2%; Score 20; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 QLTF 13  
 ||||  
 Db 1 QLTF 4

RESULT 9  
 A33995  
 adipokinetic hormone - black horse fly  
 C:Species: Tabanus atratus (black horse fly)  
 C:Date: 23-Mar-1990 #sequence\_revision 23-Mar-1990 #text\_change 31-Oct-1997  
 C:Accession: A33995  
 R:Jaffe, H.; Raina, A.K.; Riley, C.T.; Fraser, B.A.; Nachman, R.J.; Vogel, V.W.; Zhang,  
 Proc. Natl. Acad. Sci. U.S.A. 86, 8161-8164, 1989  
 A:Title: Primary structure of two neuropeptide hormones with adipokinetic and hypotrehal  
 A:Reference number: A33995; MUID:90046758  
 A:Accession: A33995  
 A:Molecule type: protein  
 A:Residues: 1-8 <JAF>  
 C:Superfamily: adipokinetic hormone  
 C:Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status predicted  
 F:8/Modified site: amidated carboxyl end (Trp) #status predicted

Query Match 31.2%; Score 20; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 QLTF 13  
 ||||  
 Db 1 QLTF 4

RESULT 10  
 S11545  
 adipokinetic hormone - nestling-sucking blowfly  
 C:Species: Protophormia terraenovae (nestling-sucking blowfly)  
 C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 31-Oct-1997  
 C:Accession: S11545

R:Gaede, G.; Wilps, H.; Kellner, R.  
 Biochem. J. 269, 309-313, 1990  
 A:Title: Isolation and structure of a novel charged member of the red-pigment-concent  
 erraenovae (Diptera).  
 A:Reference number: S11545; MUID:90351345  
 A:Accession: S11545  
 A:Molecule type: protein  
 A:Residues: 1-8 <GAE>  
 C:Superfamily: adipokinetic hormone  
 C:Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyrogluta  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
 F:8/Modified site: amidated carboxyl end (Trp) #status experimental

Query Match 31.2%; Score 20; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 QLTF 13  
 ||||  
 Db 1 QLTF 4

RESULT 11  
 A24244  
 adipokinetic hormone - bollworm  
 N:Alternate names: Hez-AKH  
 C:Species: Heliothis zea (bollworm, corn earworm, tomato fruitworm)  
 C:Date: 31-Mar-1988 #sequence\_revision 23-Mar-1995 #text\_change 31-Oct-1997  
 C:Accession: A24244  
 R:Jaffe, H.; Raina, A.K.; Riley, C.T.; Fraser, B.A.; Holman, G.M.; Wagner, R.M.; Ridg  
 Biochem. Biophys. Res. Commun. 135, 622-628, 1986  
 A:Title: Isolation and primary structure of a peptide from the corpora cardiaca of He  
 A:Reference number: A24244; MUID:86186794  
 A:Accession: A24244  
 A:Molecule type: protein  
 A:Residues: 1-9 <JAF>  
 C:Superfamily: adipokinetic hormone  
 C:Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyrogluta  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
 F:9/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 31.2%; Score 20; DB 2; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 QLTF 13  
 ||||  
 Db 1 QLTF 4

RESULT 12  
 JC1416  
 hypertrehalosemic hormone I - stick insect (Carausius morosus)  
 N:Alternate names: neuropeptide Cam-HrTH-I  
 N:Contains: hypertrehalosemic factor II  
 C:Species: Carausius morosus  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
 C:Accession: JC1416; S07157  
 R:Gaede, G.; Kellner, R.; Rinehart, K.L.; Proefke, M.L.  
 Biochem. Biophys. Res. Commun. 189, 1303-1309, 1992  
 A:Title: A tryptophan-substituted member of the AKH/RPCH family isolated from a stick  
 A:Reference number: JC1416; MUID:93129188  
 A:Accession: JC1416  
 A:Molecule type: protein  
 A:Residues: 1-10 <GAE1>  
 R:Gaede, G.; Rinehart Jr., K.L.  
 Biol. Chem. Hoppe-Seyler 368, 67-75, 1987  
 A:Title: Primary structure of the hypertrehalosaemic factor II from the corpus cardia  
 A:Reference number: S07157; MUID:87157103  
 A:Accession: S07157  
 A:Molecule type: protein

Thu Jun 6 10:09:17 2002

A:Residues: 'Z', 2-10 <GAE2>  
C:Comment: Hypertrehalosemic factor II lacks the tryptophan modification.  
C:Comment: This peptide raises hemolymph levels of trehalose in the cockroach Periplaneta  
C:Superfamily: adipokinetic hormone  
C:Keywords: amidated carboxyl end; glycoprotein; hormone; neuropeptide; pyroglutamic acid  
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
F:8/Binding site: carboxylate (Trp) (covalent) #status experimental  
F:10/Modified site: amidated carboxyl end (Thr) #status experimental

Query Match 31.2%; Score 20; DB 2; Length 10;

Best Local Similarity 100.0%; Pred. No. 1.6e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 QLTF 13

||||

Db 1 QLTF 4

RESULT 13

S09138 hypertrehalosemic hormone II - stick insect (Extatosoma tiaratum)

N/Alternate names: Cam-HrTH-II

C:Species: Extatosoma tiaratum

C:Date: 30-Jun-1992 #sequence\_revision 24-Oct-1997 #text\_change 31-Oct-1997

C:Accession: S09138

R:Gaede, G.; Rinehart, K.L.

Biol. Chem. Hoppe-Seyler 371, 345-354, 1990

A:Title: Primary structures of hypertrehalosemic neuropeptides isolated from the corpora entalis and of the stick insect Extatosoma tiaratum assigned by tandem fast atom bombardment

A:Reference number: S08995; MUID:90253659

A:Accession: S09138

A:Molecule type: protein

A:Residues: 1-10 <GAE>

A:Note: the amino-terminal residue forms pyrrolidone carboxylic acid; therefore, we have

C:Superfamily: adipokinetic hormone

C:Keywords: amidated carboxyl end; hormone; neuropeptide; pyroglutamic acid

F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

F:10/Modified site: amidated carboxyl end (Thr) #status experimental

Query Match 31.2%; Score 20; DB 2; Length 10;

Best Local Similarity 100.0%; Pred. No. 1.6e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 QLTF 13

||||

Db 1 QLTF 4

RESULT 14

A31571 hypertrehalosemic/adipokinetic hormone - bollworm

N/Alternate names: Hez-HrTH

C:Species: Heliothis zea (bollworm, corn earworm, tomato fruitworm)

C:Date: 30-Jun-1989 #sequence\_revision 23-Mar-1995 #text\_change 31-Oct-1997

C:Accession: A31571

R:Jaffe, H.; Raina, A.K.; Riley, C.T.; Fraser, B.A.; Bird, T.G.; Tseng, C.M.; Zhang, Y.S.

Biochem. Biophys. Res. Commun. 155, 344-350, 1988

A:Title: Isolation and primary structure of a neuropeptide hormone from Heliothis zea

A:Reference number: A31571; MUID:88326324

A:Accession: A31571

A:Molecule type: protein

A:Residues: 1-10 <GAE>

C:Superfamily: adipokinetic hormone

C:Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic

F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

F:10/Modified site: amidated carboxyl end (Asn) #status experimental

Query Match

Best Local Similarity 31.2%; Score 20; DB 2; Length 10;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 QLTF 13

||||

Db 1 QLTF 4

RESULT 15

B33995 hypertrehalosemic hormone - black horse fly

C:Species: Tabanus atratus (black horse fly)

C:Date: 23-Mar-1990 #sequence\_revision 23-Mar-1990 #text\_change 31-Oct-1997

C:Accession: B33995

R:Jaffe, H.; Raina, A.K.; Riley, C.T.; Fraser, B.A.; Nachman, R.J.; Vogel, V.W.; Zhang

Proc. Natl. Acad. Sci. U.S.A. 86, 8161-8164, 1989

A:Title: Primary structure of two neuropeptide hormones with adipokinetic and hypotre

A:Reference number: A33995; MUID:90046758

A:Accession: B33995

A:Molecule type: protein

A:Residues: 1-10 <GAE>

C:Superfamily: adipokinetic hormone

C:Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyrogluta

F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status predicted

F:10/Modified site: amidated carboxyl end (Tyr) #status predicted

Query Match

Best Local Similarity 31.2%; Score 20; DB 2; Length 10;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 QLTF 13

||||

Db 1 QLTF 4

Search completed: June 6, 2002, 06:23:43

Job time: 141 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:23:27 ; Search time 10.15 seconds  
(without alignments)  
49.592 Million cell updates/sec

Title: US-08-881-509-6  
Perfect score: 64  
Sequence: 1 CLVLGSGARQLTF 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 556

Minimum DB seq length: 0  
Maximum DB seq length: 13

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	35.9	11	1	ES1_RAT
2	20	31.2	8	1	AKH_TABAT
3	20	31.2	8	1	HTF2_PERAM
4	20	31.2	10	1	HTF2_CARMO
5	20	31.2	10	1	HTF_HELZE
6	20	31.2	10	1	HTF_TABAT
7	20	31.2	13	1	NP3_LYMST
8	19	29.7	8	1	CAD1_ENTFA
9	18	28.1	13	1	ITB5_BOVIN
10	18	28.1	13	1	TEMA_RANTE
11	18	28.1	13	1	TEMA_RANTE
12	17	26.6	10	1	AKHX_LOCFI
13	16	25.0	9	1	DNF1_LOCFI
14	16	25.0	10	1	MOSQ_CLYJA
15	16	25.0	12	1	CALM_TETTH
16	15	23.4	7	1	UH11_RAT
17	15	23.4	8	1	CPD1_ENTFA
18	15	23.4	8	1	HTF_TENMO
19	15	23.4	8	1	RPCH_PANBO
20	15	23.4	8	1	UPRAA_HUMAN
21	15	23.4	9	1	XYLA_STRSQ
22	15	23.4	13	1	CRBL_VESNA
23	14	21.9	8	1	ALLA_CYDPO
24	14	21.9	9	1	CONO_CONST
25	14	21.9	9	1	HUTU_KLEAE
26	14	21.9	10	1	ESTA_SCHGA
27	14	21.9	10	1	TKL2_LOCFI
28	14	21.9	10	1	URE3_MORMO
29	14	21.9	11	1	MORN_HUMAN
30	14	21.9	12	1	FREL_LITIN
31	14	21.9	13	1	CP1APLCA
32	14	21.9	13	1	HPAL_RANES
33	14	21.9	13	1	NP1_LYMST

34	14	21.9	13	1	NP2_LYMST
35	14	21.9	13	1	NP4_LYMST
36	14	21.9	13	1	NP5_LYMST
37	14	21.9	13	1	SA2A_ONCMY
38	14	21.9	13	1	SA2B_ONCMY
39	13	20.3	9	1	CCAP_CARMA
40	13	20.3	9	1	FAR9_ASCSU
41	13	20.3	9	1	LMF3_LOCFI
42	13	20.3	9	1	OXYT_BUFRE
43	13	20.3	9	1	OXYT_EISFO
44	13	20.3	10	1	TKUL_UREUN
45	13	20.3	13	1	CH60_CANEA

## ALIGNMENTS

RESULT 1					
ES1_RAT					
ID	ES1_RAT	STANDARD;	PRT;	11 AA.	
AC	P56571;				
DT	15-DEC-1998 (Rel. 37, Created)				
DT	15-DEC-1998 (Rel. 37, Last sequence update)				
DE	16-OCT-2001 (Rel. 40, Last annotation update)				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.				
OX	NCBI_TaxID=10116;				
RN	[1]				
RP	SEQUENCE.				
RC	STRAIN=WISTAR; TISSUE=Heart;				
RA	Li X.-P., Pleissner K.-P., Scheler C., Reitz-Zagrosek V., Salikov J.,				
RA	Jungblut P.R.;				
RL	Submitted (SEP-1998) to the SWISS-PROT data bank.				
CC	-!- SUBCELLULAR LOCATION: Mitochondrial (Potential).				
CC	-!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS PROTEIN				
CC	(SPOT P2) IS: 8.9, ITS MW IS: 25 kDa				
CC	-!- SIMILARITY: BELONGS TO THE ES1 FAMILY.				
KW	Mitochondrion.				
FT	NON_TER 11 11				
SQ	SEQUENCE 11 AA; 1142 MW; D862272D32C72DC2 CRC64;				

Query Match 35.9%; Score 23; DB 1; Length 11;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LVLSSS 7

Db 6 LVLST 11

RESULT 2					
AKH_TABAT					
ID	AKH_TABAT	STANDARD;	PRT;	8 AA.	
AC	P14595;				
DT	01-JAN-1990 (Rel. 13, Created)				
DT	01-FEB-1994 (Rel. 28, Last sequence update)				
DE	Adipokinet hormone (AKH) (Dipteran corpora cardiaca factor I)				
DE	(DCC I).				
OS	Tabanus atratus (Horse fly).				
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Tabanomorpha;				
OC	Tabanidae; Tabanus.				
OX	NCBI_TaxID=7207;				
RN	[1]				
RP	SEQUENCE.				
RC	TISSUE=Corpora cardiaca;				
RC	MEDLINE=90046758; PubMed=2813385;				
RA	Jaffe H., Raina A.K., Riley C.T., Fraser B.A., Nachman R.J.,				
RA	Vogel V.W., Zhang Y.-S., Hayes D.K.;				

P80179 lymnaea sta  
P80181 lymnaea sta  
P80182 lymnaea sta  
P82238 oncorhynchu  
P82239 oncorhynchu  
P38556 carcinius ma  
P43172 ascaris suu  
P41489 locusta mig  
P42995 bufo regula  
P42998 eisenia foe  
P40751 urechis uni  
P49818 canis famil

Thu Jun 6 10:09:18 2002

```

RT "Primary structure of two neuropeptide hormones with adipokinetic and
RT hypotrehalosemic activity isolated from the corpora cardiaca of horse
RT flies (Diptera).";
RL Proc. Natl. Acad. Sci. U.S.A. 86:8161-8164 (1989).
CC -!- FUNCTION: THIS HORMONE, RELEASED FROM CELLS IN THE CORPORA
CC CARDIACA AFTER THE BEGINNING OF FLIGHT, CAUSES RELEASE OF
CC DIGLYCERIDES FROM THE FAT BODY AND THEN STIMULATES THE FLIGHT
CC MUSCLES TO USE THESE DIGLYCERIDES AS AN ENERGY SOURCE.
CC -!- SIMILARITY: BELONGS TO THE AKH / HRFH / RPCH FAMILY.
CC PIR: A33995; A33995.
CC InterPro: IPR002047; AKH.
CC PROSITE: PS00256; AKH; 1.
DR Neuropeptide; Amidation; Flight.
KW Neuropeptide; 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 8 8 AMIDATION.
FT SEQUENCE 8 AA; 949 MW; 86786771A9D1A736 CRC64;
SQ

Query Match 31.2%; Score 20; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 QLTF 13
DB 1 QLTF 4

RESULT 3
HTF2_PERAM STANDARD; PRT; 8 AA.
AC P04549;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Hypertrehalosemic factor II (Neuropeptide M-II) (Periplanetin CC-2)
DE (Pea-CAH-II) (Leb-CC-II) (Hypertrehalosemic neuropeptide II).
OS Periplaneta americana (American cockroach),
OS Leptinotarsa decemlineata (Colorado potato beetle), and
OS Blatta orientalis (Oriental cockroach).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC Blattodea; Blattidae; Periplaneta.
OX NCBI_TaxID=6978, 7539, 6976;
RN [1]
RP SEQUENCE.
RC SPECIES=P.americana;
RX MEDLINE=85046530; PubMed=6548628;
RA Witten J.L., Schaffer M.H., O'Shea M., Cook J.C., Hemling M.E.,
RA Rinehart K.L. Jr.;
RT "Structures of two cockroach neuropeptides assigned by fast atom
RT bombardment mass spectrometry.";
RL Biochem. Biophys. Res. Commun. 124:350-358 (1984).
RN [2]
RP SEQUENCE.
RC SPECIES=P.americana;
RX MEDLINE=84298179; PubMed=6591205;
RA Scarborough R.M., Jamieson G.C., Kalish F., Kramer S.J., McEnroe G.A.,
RA Miller C.A., Schooley D.A.;
RT "Isolation and primary structure of two peptides with
RT cardioacceleratory and hyperglycemic activity from the corpora
RT cardiaca of Periplaneta americana.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:5575-5579 (1984).
RN [3]
RP SEQUENCE.
RC SPECIES=L.decemlineata; TISSUE=Corpora cardiaca;
RX MEDLINE=90160053; PubMed=2576128;
RA Gaede G., Kellner R.;
RT "The metabolic neuropeptides of the corpus cardiacum from the potato
RT beetle and the American cockroach are identical.";
RL Peptides 10:1287-1289 (1989).
RN [4]
RP SEQUENCE.
RC SPECIES=B.orientalis; TISSUE=Corpora cardiaca;

```

```

RX MEDLINE=90253659; PubMed=2340112;
RA Gaede G., Rinehart K.L. Jr.;
RT "Primary structures of hypertrehalosemic neuropeptides isolated from
RT the corpora cardiaca of the cockroaches Leucophaea maderae,
RT Gromphadorhina portentosa, Blattella germanica and Blatta orientalis
RT and of the stick insect Extatosoma tiaratum assigned by tandem fast
RT atom bombardment mass spectrometry.";
RL Biol. Chem. Hoppe-Seyler 371:345-354 (1990).
CC -!- FUNCTION: HYPERTREHALOSEMIC FACTORS ARE NEUROPEPTIDES THAT
CC ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS
CC THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).
CC -!- SIMILARITY: BELONGS TO THE AKH / HRFH / RPCH FAMILY.
CC PIR: A05170; A05170.
DR PIR: S08996; S08996.
DR PIR: B44960; B44960.
DR PIR: B49823; B49823.
DR InterPro: IPR002047; AKH.
DR PROSITE: PS00256; AKH; 1.
KW Neuropeptide; Amidation.
FT MOD_RES 1 1
FT MOD_RES 8 8 PYRROLIDONE CARBOXYLIC ACID.
FT SEQUENCE 8 AA; 1006 MW; 86745771A9D1A736 CRC64;
SQ

Query Match 31.2%; Score 20; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 QLTF 13
DB 1 QLTF 4

RESULT 4
HTF2_CARMO STANDARD; PRT; 10 AA.
AC P11385;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypertrehalosemic factor II (HFF-II) (HRTH-II) (Hypertrehalosemic
DE neuropeptide II).
OS Carausius morosus (Indian stick insect), and
OS Extatosoma tiaratum (Stick insect).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Phasmatodea; Heteronemiidae;
OC Carausius
OX NCBI_TaxID=7022, 7024;
RN [1]
RP SEQUENCE.
RC SPECIES=C.morosus; TISSUE=Corpora cardiaca;
RX MEDLINE=87157103; PubMed=3828078;
RA Gaede G., Rinehart K.L. Jr.;
RT "Primary structure of the hypertrehalosemic factor II from the
RT corpus cardiacum of the Indian stick insect, Carausius morosus,
RT determined by fast atom bombardment mass spectrometry.";
RL Biol. Chem. Hoppe-Seyler 368:67-75 (1987).
RN [2]
RP SEQUENCE.
RC SPECIES=E.tiaratum; TISSUE=Corpora cardiaca;
RX MEDLINE=90253659; PubMed=2340112;
RA Gaede G., Rinehart K.L. Jr.;
RT "Primary structures of hypertrehalosemic neuropeptides isolated from
RT the corpora cardiaca of the cockroaches Leucophaea maderae,
RT Gromphadorhina portentosa, Blattella germanica and Blatta orientalis
RT and of the stick insect Extatosoma tiaratum assigned by tandem fast
RT atom bombardment mass spectrometry.";
RL Biol. Chem. Hoppe-Seyler 371:345-354 (1990).
RN [3]
RP CARBOHYDRATE-LINKAGE SITE.
RC SPECIES=C.morosus; TISSUE=Corpora cardiaca;
RX MEDLINE=93129188; PubMed=1482345;
RA Gaede G., Kellner R., Rinehart K.L. Jr., Proefke M.L.;

```

RT "A" tryptophan-substituted member of the AKH/RPCH family isolated from  
 RL a stick insect corpus cardiacum."  
 CC Biochem. Biophys. Res. Commun. 189:1303-1309(1992).  
 CC -1- FUNCTION: HYPERTREHALOSAEMIC FACTORS ARE NEUROPEPTIDES THAT  
 CC ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS  
 CC THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).  
 CC -1- MASS SPECTROMETRY: MW=1308.61; METHOD-FAB.  
 CC -1- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.  
 DR PIR: S07157; S07157.  
 DR PIR: S09138; S09138.  
 DR InterPro: IPR002047; AKH.  
 DR PROSITE: PS00256; AKH; 1.  
 KW Neuropeptide; Amidation; Glycoprotein.  
 FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 FT CARBOHYD 8 8 C-LINKED (MAN) (PROBABLE).  
 FT MOD\_RES 10 10 AMIDATION.  
 FT MOD\_RES 10 10  
 SQ SEQUENCE 10 AA; 1164 MW; 9B9036745771A9D1 CRC64;

Query Match 31.2%; Score 20; DB 1; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 5.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 QLTF 13  
 Db 1 QLTF 4

RESULT 5  
 HTF\_HELZE  
 ID HTF\_HELZE STANDARD; PRT; 10 AA.  
 AC P16353;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Hypertrehalosemic hormone (Hez-HRTH).  
 OS Heliothis zea (Corn earworm) (Bollworm).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Noctuidae; Noctuidae; Heliothinae; Helicoverpa.  
 OX NCBI\_TaxID=7113;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Corpora cardiaca;  
 RX MEDLINE=88326324; PubMed=3415690;  
 RA Jaffe H., Raina A.K., Riley C.T., Fraser B.A., Bird T.G.,  
 RA Tseng C.M., Zhang Y.S., Hayes D.K.;  
 RT "Isolation and primary structure of a neuropeptide hormone from  
 RT Heliothis zea with hypertrehalosemic and adipokinetic activities.";  
 RL Biochem. Biophys. Res. Commun. 155:344-350(1988).  
 CC -1- FUNCTION: HYPERTREHALOSAEMIC FACTORS ARE NEUROPEPTIDES THAT  
 CC ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS  
 CC THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).  
 CC -1- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.  
 DR PIR: A31571; A31571.  
 DR InterPro: IPR002047; AKH.  
 DR PROSITE: PS00256; AKH; 1.  
 KW Neuropeptide; Amidation.  
 FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 FT MOD\_RES 10 10 AMIDATION.  
 FT MOD\_RES 10 10  
 SQ SEQUENCE 10 AA; 1096 MW; 8E70367865A5B9D1 CRC64;

Query Match 31.2%; Score 20; DB 1; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 5.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 QLTF 13  
 Db 1 QLTF 4

RESULT 6

HTF\_TABAT  
 ID HTF\_TABAT STANDARD; PRT; 10 AA.  
 AC P14596;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Hypertrehalosemic factor (HOTH) (Dipteran corpora cardiaca factor II)  
 DE (DCC II).  
 OS Tabanus atratus (Horse fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Tabanomorpha;  
 OC Tabanidae; Tabanus.  
 OX NCBI\_TaxID=7207;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Corpora cardiaca;  
 RX MEDLINE=90046758; PubMed=2813385;  
 RA Jaffe H., Raina A.K., Riley C.T., Fraser B.A., Nachman R.J.,  
 RA Vogel V.W., Zhang Y.S., Hayes D.K.;  
 RT "Primary structure of two neuropeptide hormones with adipokinetic and  
 RT hypotrehalosemic activity isolated from the corpora cardiaca of horse  
 RT flies (Diptera)."  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:8161-8164(1989).  
 CC -1- FUNCTION: HYPERTREHALOSAEMIC FACTORS ARE NEUROPEPTIDES THAT  
 CC ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS  
 CC THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).  
 CC -1- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.  
 DR PIR: B33995; B33995.  
 DR InterPro: IPR002047; AKH.  
 DR PROSITE: PS00256; AKH; 1.  
 KW Neuropeptide; Amidation.  
 FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 FT MOD\_RES 10 10 AMIDATION.  
 FT MOD\_RES 10 10  
 SQ SEQUENCE 10 AA; 1169 MW; 916036786771A9D1 CRC64;

Query Match 31.2%; Score 20; DB 1; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 5.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 QLTF 13  
 Db 1 QLTF 4

RESULT 7  
 NP3\_LYMST  
 ID NP3\_LYMST STANDARD; PRT; 13 AA.  
 AC P80180;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-JUL-1993 (Rel. 26, Last annotation update)  
 DE Lymnaea-Df-amide 3.  
 DE Lymnaea stagnalis (Great pond snail).  
 OS Eukaryota; Metazoa; Mollusca; Gastropoda; Basommatophora;  
 OC Lymnaeidae; Lymnaea.  
 OX NCBI\_TaxID=6523;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Ganglion;  
 RX MEDLINE=93238777; PubMed=8477756;  
 RA Johnsen A.H., Rehfeld J.F.;  
 RT "Lymnaeidae, a new family of neuropeptides from the pond snail,  
 RT Lymnaea stagnalis. Clue to cholecystokinin immunoreactivity in  
 RT invertebrates?";  
 RL Eur. J. Biochem. 213:875-879(1993).  
 CC -1- SIMILARITY: RELATED TO THE CHOLECYSTOKININ (CCK) FAMILY.  
 DR PIR: S32473; S32473.  
 KW Neuropeptide; Amidation.  
 FT MOD\_RES 13 13  
 FT MOD\_RES 12 12 AMIDATION.  
 FT UNSURE 12 12  
 SQ SEQUENCE 13 AA; 1462 MW; 9CA07BA3F5D5B865 CRC64;

Thu Jun 6 10:09:18 2002

DR InterPro: IPR001169; Integrin\_beta.C.  
 DR PROSITE: PS00243; INTEGRIN\_BETA; PARTIAL.  
 KW Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;  
 KW Repeat.  
 FT NON\_TER 13 13  
 SQ SEQUENCE 13 AA; 1299 MW; 844197D005B9B865 CRC64;

Query Match 31.2%; Score 20; DB 1; Length 13;  
 Best Local Similarity 80.0%; Pred. No. 7.2e+02;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 LSGSA 8

Db 5 ISGSA 9

RESULT 8

CAD1\_ENTFA STANDARD; PRT; 8 AA.  
 AC P13268;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-FEB-1991 (Rel. 17, Last annotation update)  
 DE Sex pheromone CAD1.  
 OS Enterococcus faecalis (Streptococcus faecalis).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;  
 OC Enterococcus.  
 OX NCBI\_TaxID=1351;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=85051899; PubMed=6437872;  
 RA Mori M., Sakakami Y., Narita M., Isogai A., Fujino M., Kitada C.,  
 RA Craig R.A., Clewell D.B., Suzuki A.;  
 RA "Isolation and structure of the bacterial sex pheromone, cad1, that  
 RT induces plasmid transfer in Streptococcus faecalis.";  
 RL FEBS Lett. 178:97-100(1984).  
 CC -1- FUNCTION: CAD1 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE  
 CC HEMOLYSIN PLASMID PAD1.  
 KW Pheromone.  
 SQ SEQUENCE 8 AA; 819 MW; 047DD732C735B9C7 CRC64;

Query Match 29.7%; Score 19; DB 1; Length 8;  
 Best Local Similarity 80.0%; Pred. No. 1e+05;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVLGS 6

Db 4 LVLGS 8

RESULT 9

ITB5\_BOVIN STANDARD; PRT; 13 AA.  
 AC P80747;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Integrin beta-5 (fragment).  
 GN ITGB5.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE.  
 RX TISSUE=Mammary gland;  
 RX MEDLINE=97299777; PubMed=9154926;  
 RA Andersen M.H., Berglund L., Rasmussen J.T., Petersen T.E.;  
 RA "Bovine PAS-6/7 binds alpha v beta 5 integrins and anionic  
 RT phospholipids through two domains.";  
 RL Biochemistry 36:5441-5446(1997).  
 CC -1- FUNCTION: INTEGRIN ALPHA-V/BETA-5 IS A RECEPTOR FOR FIBRONECTIN.  
 CC IT RECOGNIZES THE SEQUENCE R-G-D IT ITS LIGAND.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-5  
 CC ASSOCIATES WITH ALPHA-V.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.

Query Match 28.1%; Score 18; DB 1; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SGSA 8

Db 6 SGSA 9

RESULT 10

TEMA\_RANTE STANDARD; PRT; 13 AA.  
 AC P56917;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Temporin A.  
 OS Rana temporaria (European common frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.  
 OX NCBI\_TaxID=8407;  
 RN [1]  
 RP SEQUENCE, AND SYNTHESIS.  
 RC TISSUE=Skin;  
 RX MEDLINE=97175050; PubMed=9022710;  
 RA Simmaco M., Mignogna G., Canofeni S., Miele R., Mangoni M.L.,  
 RA Barra D.;  
 RA "Temporins, antimicrobial peptides from the European red frog Rana  
 RT temporaria.";  
 RL Eur. J. Biochem. 242:788-792(1996).  
 CC -1- FUNCTION: HAS ANTIBACTERIAL ACTIVITY AGAINST GRAM-POSITIVE  
 CC BACTERIA.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE BREVININ/ESCULENTIN/GAegurin/RUGOSIN  
 CC FAMILY.  
 KW Amphibian skin; Antibiotic; Amidation; Multigene family.  
 FT MOD\_RES 13 13  
 SQ SEQUENCE 13 AA; 1398 MW; 2653612B9DECD408 CRC64;

Query Match 28.1%; Score 18; DB 1; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VLGS 6

Db 8 VLGS 11

RESULT 11

TEMA\_RANTE STANDARD; PRT; 13 AA.  
 AC P56921;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Temporin F.  
 OS Rana temporaria (European common frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.  
 OX NCBI\_TaxID=8407;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Skin;  
 RX MEDLINE=97175050; PubMed=9022710;

RA Simmaco M., Mignogna G., Canofeni S., Miele R., Mangoni M.L.,  
 RA Barra D.;  
 RT "Temporins, antimicrobial peptides from the European red frog Rana  
 RL Eur. J. Biochem. 242:788-792(1996).  
 CC -1- FUNCTION: HAS ANTI-BACTERIAL ACTIVITY AGAINST GRAM-NEGATIVE AND  
 CC GRAM-POSITIVE BACTERIA.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE BREVININ/ESCULENTIN/GAECURIN/RUGOSIN  
 CC FAMILY.  
 KW Amphibian skin; Antibiotic; Amidation; Multigene family.  
 FT MOD\_RES 13 13  
 SQ SEQUENCE 13 AA; 1370 MW; 2653612B9DECC338 CRC64;

Query Match 28.1%; Score 18; DB 1; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VLSG 6  
 ||||  
 DB 8 VLSG 11

## RESULT 12

AKHX\_LOCM1 STANDARD; PRT; 10 AA.  
 ID AKHX\_LOCM1  
 AC P81626;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Peptide hormone.  
 OS Locusta migratoria (Migratory locust).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;  
 OC Acridoidea; Acridoidea; Acrididae; Locusta.  
 RN NCBI\_TaxID=7004;  
 RP SEQUENCE.  
 RC TISSUE=Corpora cardiaca;  
 RA Siebert K.J.;  
 RL Submitted (DEC-1998) to the SWISS-PROT data bank.  
 CC -1- FUNCTION: PROBABLY INVOLVED IN THE REGULATION OF LOCUST  
 CC INTERMEDIARY METABOLISM, BEHAVIOR AND/OR DEVELOPMENT.  
 CC -1- SIMILARITY: SOME SIMILARITY TO THE AKH / HRTH / RPCH FAMILY.  
 DR InterPro: IPR002047; AKH.  
 KW Neuropeptide; Amidation.  
 FT MOD\_RES 1 1  
 FT MOD\_RES 10 10  
 SQ SEQUENCE 10 AA; 1222 MW; 81BFF67AB415B9D1 CRC64;

Query Match 26.6%; Score 17; DB 1; Length 10;  
 Best Local Similarity 75.0%; Pred. No. 2.1e+03;  
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 10 QLTF 13  
 ||||  
 DB 1 QVTF 4

## RESULT 13

DNFL\_LOCM1 STANDARD; PRT; 9 AA.  
 ID DNFL\_LOCM1  
 AC P16339;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Locupressin (Diuretic neuropeptide F1/F2).  
 OS Locusta migratoria (Migratory locust).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;  
 OC Acridoidea; Acridoidea; Acrididae; Locusta.

## OX NCBI\_TaxID=7004;

RN SEQUENCE.  
 RP TISSUE=Suboesophageal ganglion, and Thoracic ganglion;  
 RC MEDLINE=88077077; PubMed=3689410;  
 RA Proux J.P., Miller C.A., Li J.P., Carney R.L., Girardie A.,  
 RA Delage M., Schooley D.A.;  
 RT "Identification of an arginine vasopressin-like diuretic hormone from  
 RT Locusta migratoria.";  
 RL Biochem. Biophys. Res. Commun. 149:180-186(1987).  
 CC -1- FUNCTION: DIURETIC HORMONE.  
 CC -1- SUBUNIT: F2 IS AN ANTIPARALLEL DISULFIDE LINKED DIMER OF F1.  
 CC -1- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.  
 DR PIR: A29477; A29477.  
 DR InterPro: IPR000981; Neurohypophys\_horm.  
 DR Pfam: PF00220; hormone4; 1.  
 DR PROSITE: PS00264; NEUROHYPOPHYS\_HORM; 1.  
 KW Hormone; Neuropeptide; Amidation.  
 FT DISULFID 1 6  
 FT DISULFID 1 1  
 FT DISULFID 6 6  
 FT DISULFID 9 9  
 FT MOD\_RES 9 9  
 SQ SEQUENCE 9 AA; 976 MW; 56EB176EB451A057 CRC64;

Query Match 25.0%; Score 16; DB 1; Length 9;  
 Best Local Similarity 66.7%; Pred. No. 1e+05;  
 Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLV 3  
 ||:  
 DB 1 CLI 3

## RESULT 14

MOSQ\_CLYJA STANDARD; PRT; 10 AA.  
 ID MOSQ\_CLYJA  
 AC P19962;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-FEB-1991 (Rel. 17, Last annotation update)  
 DE [Gln-6]-mosact.  
 OS Clypeaster japonicus (Sand dollar).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinoidea; Echinoidea; Gnathostomata; Clypeasteroidea;  
 OC Clypeasteridae; Clypeaster.  
 RN NCBI\_TaxID=7644;  
 RP SEQUENCE.  
 RC TISSUE=Egg jelly;  
 RA Suzuki N., Kurita M., Yoshino K., Kajitara H., Nomura K., Yamaguchi M.;  
 RT "Purification and structure of mosact and its derivatives from the  
 RT egg jelly of the sea urchin Clypeaster japonicus.";  
 RL Zool. Sci. 4:649-656(1987).  
 CC -1- FUNCTION: STIMULATES SPERM RESPIRATION AND MOTILITY.  
 DR PIR: JN0025; JN0025.  
 SQ SEQUENCE 10 AA; 1019 MW; 9AFB032456DDC5BA CRC64;

Query Match 25.0%; Score 16; DB 1; Length 10;  
 Best Local Similarity 57.1%; Pred. No. 3.3e+03;  
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 SGSARQL 11  
 ||||  
 DB 2 SDSAQL 8

## RESULT 15

CALM\_TETH STANDARD; PRT; 12 AA.  
 ID CALM\_TETH  
 AC Q05055;  
 DT 01-FEB-1994 (Rel. 28, Created)

Thu Jun 6 10:09:18 2002

01-FEB-1994 (Rel. 28, Last sequence update)  
01-MAR-2002 (Rel. 41, Last annotation update)  
Calmodulin (Fragment).  
Tetrahymena thermophila.  
Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;  
Tetrahymenina; Tetrahymena.  
NCBI\_TaxID=5911;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE=93281388; PubMed=8506136;  
Kato M., Hirono M., Takemasa T., Kimura M., Watanabe Y.;  
"A micro-nucleus-specific sequence exists in the 5'-upstream region of  
calmodulin gene in Tetrahymena thermophila.";  
Nucleic Acids Res. 21:2409-2414(1993).  
-1- FUNCTION: CALMODULIN MEDIATES THE CONTROL OF A LARGE NUMBER OF  
ENZYMES BY CA(++) . AMONG THE ENZYMES TO BE STIMULATED BY THE  
CALMODULIN-CA(++) COMPLEX ARE A NUMBER OF PROTEIN KINASES AND  
PHOSPHATASES.  
-1- MISCELLANEOUS: THIS PROTEIN HAS FOUR FUNCTIONAL CALCIUM-BINDING  
SITES.  
-1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; D12774; BAA02239.1; .  
HSP; P02593; 2CLN.  
InterPro; IPR002048; EF-hand.  
PROSITE; PS00018; EF\_HAND; PARTIAL.  
Calcium-binding; Repeat; Acetylation.  
INIT\_MET 0 0 BY SIMILARITY.  
MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
NON\_TER 12 12  
SEQUENCE 12 AA; 1393 MW; 83F31CD443DB1B01 CRC64;  
  
Query Match 25.0%; Score 16; DB 1; Length 12;  
Best Local Similarity 80.0%; Pred. No. 4e+03;  
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
OY 8 ARQLT 12  
Db 1 ADOLT 5

Search completed: June 6, 2002, 06:26:36  
Job time: 189 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:23:07 ; Search time 23.81 Seconds  
(without alignments)  
94.453 Million cell updates/sec

Title: US-08-881-509-6

Perfect score: 64

Sequence: 1 CLVLSGSARQLTF 13

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 2182

Minimum DB seq length: 0

Maximum DB seq length: 13

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL19:\*  
1: sp-archaea:\*  
2: sp-bacteria:\*  
3: sp-fungi:\*  
4: sp-human:\*  
5: sp-invertebrate:\*  
6: sp-mammal:\*  
7: sp-mhc:\*  
8: sp-organelle:\*  
9: sp-phage:\*  
10: sp-plant:\*  
11: sp-rodent:\*  
12: sp-virus:\*  
13: sp-vertebrate:\*  
14: sp-unclassified:\*  
15: sp-virus:\*  
16: sp-bacteriaph:\*  
17: sp-archaeap:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	32.8	10	5 Q25356	Q25356 locusta mig
2	21	32.8	11	11 Q9R0K9	Q9R0K9 mus musculu
3	21	32.8	12	10 Q945C3	Q945C3 crypthecodi
4	19	29.7	13	4 Q9UM84	Q9UM84 homo sapien
5	19	29.7	13	5 O61340	O61340 panulirus i
6	18	28.1	8	2 O09258	O09258 synchococ
7	18	28.1	10	4 Q9P219	Q9P219 homo sapien
8	18	28.1	11	11 P97330	P97330 mus musculu
9	18	28.1	12	11 Q64296	Q64296 mus musculu
10	18	28.1	13	6 Q9TQ51	Q9TQ51 equus cabal
11	18	28.1	13	13 Q9PWP4	Q9PWP4 dissostichu
12	17	26.6	9	13 Q92009	Q92009 gallus gall
13	17	26.6	11	15 Q98YS3	Q98YS3 human immun
14	17	26.6	12	2 P83054	P83054 bacteroides
15	17	26.6	12	6 Q9TQY4	Q9TQY4 bos taurus
16	17	26.6	13	6 Q9TUD7	Q9TUD7 bos taurus

17	17	26.6	13	10 Q9XIW6	Q9XIW6 oryza sativ
18	17	26.6	13	11 Q62352	Q62352 mus musculu
19	17	26.6	13	11 Q62354	Q62354 mus musculu
20	17	26.6	13	11 Q62355	Q62355 mus musculu
21	16	25.0	8	4 Q96RN9	Q96RN9 homo sapien
22	16	25.0	8	12 Q89498	Q89498 murine hepa
23	16	25.0	8	12 Q83349	Q83349 murine coro
24	16	25.0	9	11 Q35953	Q35953 mus musculu
25	16	25.0	9	11 Q9PYK1	Q9PYK1 simian viru
26	16	25.0	9	12 Q9IBM8	Q9IBM8 simian viru
27	16	25.0	10	2 Q9X3M2	Q9X3M2 prochloroco
28	16	25.0	10	8 Q37103	Q37103 saccharomyc
29	16	25.0	11	5 P82700	P82700 leucophaea
30	16	25.0	12	10 Q945C2	Q945C2 crypthecodi
31	16	25.0	12	12 O10421	O10421 influenza a
32	16	25.0	13	8 Q94RE2	Q94RE2 leptomonas
33	16	25.0	13	10 Q9S8N1	Q9S8N1 hordeum vul
34	16	25.0	13	11 P97944	P97944 mus musculu
35	16	25.0	13	11 Q9QVDS	Q9QVDS rattus sp.
36	16	25.0	13	12 Q9WMG5	Q9WMG5 sigma virus
37	15	23.4	7	2 P70804	P70804 azotobacter
38	15	23.4	8	6 P82929	P82929 bos taurus
39	15	23.4	8	11 Q60615	Q60615 mus musculu
40	15	23.4	8	12 Q89965	Q89965 polyomaviru
41	15	23.4	8	13 P82079	P82079 limnodynast
42	15	23.4	9	5 Q27396	Q27396 babesia bov
43	15	23.4	9	6 Q9TRSO	Q9TRSO oryctolagus
44	15	23.4	10	10 Q08938	Q08938 nicotiana t
45	15	23.4	10	10 Q94119	Q94119 zea mays (m

## ALIGNMENTS

RESULT 1  
Q25356 PRELIMINARY; PRT; 10 AA.  
ID Q25356  
AC Q25356;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DE VITELLOGENIN B (FRAGMENT).  
OS Locusta migratoria (Migratory locust).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;  
OC Acridomorpha; Acridoidea; Acrididae; Locusta.  
OX NCBI\_TaxID=7004;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88003971; PubMed=2820677;  
RA Locke J., White B.N., Wyatt G.R.;  
RT "Cloning and 5' end nucleotide sequences of two juvenile hormone-  
inducible vitellogenin genes of the African migratory locust.";  
RL DNA 6:331-342(1987).  
DR EMBL: M17334; AAA29285.1; -.  
FT NON\_TER  
SQ SEQUENCE 10 AA; 1116 MW; DIB31165B720572D CRC64;

Query Match 32.8%; Score 21; DB 5; Length 10;  
Best Local Similarity 80.0%; Pred. No. 1.8e+03;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LVLSS 6  
Db 4 LILSS 8

RESULT 2  
Q9R0K9 PRELIMINARY; PRT; 11 AA.  
ID Q9R0K9  
AC Q9R0K9;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)

Thu Jun 6 10:09:20 2002

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE BCL-2 ASSOCIATED X PROTEIN (FRAGMENT).  
 GN BAX OR BAX.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20035745; PubMed=10570968;  
 RA Igata E., Inoue T., Ohtani-Fujita N., Sowa Y., Tsujimoto Y., Sakai T.;  
 RT "Molecular cloning and functional analysis of the murine bax gene  
 RT promoter."; 415(1999).  
 RL Gene 238:407-415(1999).  
 DR EMBL; AB029557; BAA82406.1; -.  
 DR MGD; MGI:99702; Bax.  
 FT NON\_TER 11 11  
 FT SEQUENCE 11 AA; 1037 MW; 90AAB34E36DB1865 CRC64;  
 SQ SEQUENCE 11 AA; 1037 MW; 90AAB34E36DB1865 CRC64;  
 Query Match 32.8%; Score 21; DB 11; Length 11;  
 Best Local Similarity 50.0%; Pred. No. 2e+03;  
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 LSGSARQL 11  
 Db : ||| |  
 1 MDGSGEQL 8  
 RESULT 3  
 ID Q945C3 PRELIMINARY; PRT; 12 AA.  
 AC Q945C3;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE P80 PROTEIN (FRAGMENT).  
 OS Cryptocodium cohnii (Dinoflagellate).  
 OC Eukaryota; Alveolata; Dinophyceae; Gonyaulacales; Cryptophyceae;  
 OC Cryptocodium.  
 OX NCBI\_TaxID=2866;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99433238; PubMed=10505419;  
 RA Ausseil J., Soyer-Gobillard M.O., Geraud M.L., Bhaud Y., Baines I.,  
 RA Preston T., Moreau H.;  
 RT "Characterization of p80, a novel nuclear and cytoplasmic protein in  
 RT dinoflagellates."; 150:197-211(1999).  
 RL Protist 150:197-211(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Guillebaud D., Derelle E., Lozano J.C., Bingham S., Moreau H.;  
 RT "A single TBP-like protein is present in the marine unicellular  
 RT organism: the dinoflagellate cryptocodium cohnii."; 150:197-211(1999).  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF417568; AAL15906.1; -.  
 FT NON\_TER 12 12  
 FT SEQUENCE 12 AA; 1364 MW; 615BF873FE204414 CRC64;  
 SQ SEQUENCE 12 AA; 1364 MW; 615BF873FE204414 CRC64;  
 Query Match 32.8%; Score 21; DB 10; Length 12;  
 Best Local Similarity 57.1%; Pred. No. 2.2e+03;  
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 7 SARQLTF 13  
 Db ||||| |  
 3 SARRILF 9  
 RESULT 4  
 ID Q9UM84 PRELIMINARY; PRT; 13 AA.  
 AC Q9UM84  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE RET-ELE1 PROTEIN (FRAGMENT).  
 GN RET-ELE1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96400323; PubMed=8806699;  
 RA Fugazzola L., Pierotti M.A., Viganò E., Pacini F., Vorontsova T.V.,  
 RA Bougarzone L.;  
 RT "Molecular and biochemical analysis of RET/PTC4, a novel oncogenic  
 RT rearrangement between RET and ELE1 genes, in a post-Chernobyl  
 RT papillary thyroid cancer."; 13:1093-1097(1996).  
 RL Oncogene 13:1093-1097(1996).  
 DR EMBL; S83049; AAD14423.1; -.  
 FT NON\_TER 1 1  
 FT SEQUENCE 13 AA; 1463 MW; 7C84ED9161C1A69B CRC64;  
 SQ SEQUENCE 13 AA; 1463 MW; 7C84ED9161C1A69B CRC64;  
 Query Match 29.7%; Score 19; DB 4; Length 13;  
 Best Local Similarity 42.9%; Pred. No. 5.8e+03;  
 Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 7 SARQLTF 13  
 Db | : : | |  
 3 SSAEMTF 9  
 RESULT 5  
 ID O61340 PRELIMINARY; PRT; 13 AA.  
 AC O61340;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE POTASSIUM CHANNEL (FRAGMENT).  
 OS Panulirus interruptus (California spiny lobster).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Palinura;  
 OC Palinuroidea; Palinuridae; Panuliridae; Panulirus.  
 OX NCBI\_TaxID=6735;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98330950; PubMed=9666521;  
 RA Kim M., Baro D.J., Lanning C.C., Doshi M., Moskowitz H.S., Farnham J.,  
 RA Harris-Warrick R.M.;  
 RT "Expression of Panulirus shaker potassium channel splice variants."; 5:291-304(1998).  
 RL Recept. Channels 5:291-304(1998).  
 DR EMBL; AF017135; AAC05915.1; -.  
 FT NON\_TER 13 13  
 FT SEQUENCE 13 AA; 1336 MW; CBA864F1E31E31AD CRC64;  
 SQ SEQUENCE 13 AA; 1336 MW; CBA864F1E31E31AD CRC64;  
 Query Match 29.7%; Score 19; DB 5; Length 13;  
 Best Local Similarity 80.0%; Pred. No. 5.8e+03;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 LVLSG 6  
 Db | : | | |  
 8 LLLSG 12  
 RESULT 6  
 ID O09258 PRELIMINARY; PRT; 8 AA.  
 AC O09258  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)



DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)  
 DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 GN NIFH (FRAGMENT).  
 OS Synchococcus sp. (strain PCC 8801 / RF-1) (Cyanothecae PCC 8801).  
 OC Bacteria; Cyanobacteria; Chroococcales; Cyanothecae.  
 OX NCBI\_TaxID=41431;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RF-1;  
 RA MEDLINE=99231861; PubMed=10217509;  
 RA Huang T.C., Lin R.F., Chu M.K., Chen H.M.;  
 RT "Organization and expression of nitrogen-fixation genes in the aerobic  
 RT nitrogen-fixing unicellular cyanobacterium Synchococcus sp. strain  
 RT RF-1";  
 RL Microbiology 145:743-753(1999).  
 DR EMBL; AF001780; AAC33369.1; -;  
 FT NON\_TER 8  
 SQ SEQUENCE 8 AA; 985 MW; F16B59CDD046C406 CRC64;

Query Match 28.1%; Score 18; DB 2; Length 8;  
 Best Local Similarity 60.0%; Pred. No. 5.6e+05;  
 Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 9 RQLTF 13  
 Db [1]  
 2 RQIAF 6

RESULT 7  
 QP229 ID QP229 PRELIMINARY; PRT; 10 AA.  
 AC QP229;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DE EPIDERMAL GROWTH FACTOR RECEPTOR (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91337091; PubMed=1678600;  
 RA Humphrey P.A., Gangarosa L.M., Wong A.J., Archer G.E.,  
 RA Lund-Johansen M., Bjerkvig R., Laerum O.D., Friedman H.S.,  
 RA Bigner D.D.;  
 RT "Deletion-mutant epidermal growth factor receptor in human gliomas:  
 RT effects of type II mutation on receptor function.";  
 RL Biochem. Biophys. Res. Commun. 178:1413-1420(1991).  
 DR EMBL; S51343; AAB19486.2; -;  
 KW Receptor.  
 FT NON\_TER 1  
 FT NON\_TER 10  
 SQ SEQUENCE 10 AA; 1020 MW; 435DEAE8B7B1727 CRC64;

Query Match 28.1%; Score 18; DB 4; Length 10;  
 Best Local Similarity 50.0%; Pred. No. 6.8e+03;  
 Matches 3; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CLVLSG 6  
 Db [1]  
 1 CKLLEG 6

RESULT 8  
 P97330 ID P97330 PRELIMINARY; PRT; 11 AA.  
 AC P97330;  
 DT 01-MAY-1997 (Tremblrel. 03, Created)  
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE CD45-AP (LSM-1).  
 GN PTPRCAP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=97124850; PubMed=8954783;  
 RA Bruyns E., Mincheva A., Bruyns R.M., Kirchgessner H., Weitz S.,  
 RA Lichter P., Meier S., Schraven B.;  
 RT "Sequence, genomic organization and chromosomal localization of the  
 RT human LPAAP (PTPRCAP) and Mouse CD45-AP/LSM-1 genes";  
 RL Genomics 38:79-83(1996).  
 DR EMBL; X97268; CAA65923.1; -;  
 DR MGD; MGI:97811; Ptpcap.  
 SQ SEQUENCE 11 AA; 1150 MW; 50695413B5A772C7 CRC64;

Query Match 28.1%; Score 18; DB 11; Length 11;  
 Best Local Similarity 80.0%; Pred. No. 7.5e+03;  
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CLVLS 5  
 Db [1]  
 3 CLVPS 7

RESULT 9  
 Q64296 ID Q64296 PRELIMINARY; PRT; 12 AA.  
 AC Q64296;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
 DE CD44 ANTIGEN (FRAGMENT).  
 GN CD44.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SWISS WEBSTER;  
 RA Yu Q., Toole B.;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SWISS WEBSTER;  
 RA Yu Q., Toole B.P.;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U57612; AAC52806.1; -;  
 DR EMBL; U57611; AAB08756.1; -;  
 DR MGD; MGI:88338; Cd44.  
 FT NON\_TER 1  
 SQ SEQUENCE 12 AA; 1235 MW; CBFDEB7444ADC2D2 CRC64;

Query Match 28.1%; Score 18; DB 11; Length 12;  
 Best Local Similarity 75.0%; Pred. No. 8.2e+03;  
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CLVL 4  
 Db [1]  
 2 CLVV 5

RESULT 10  
 Q9TOS1 ID Q9TOS1 PRELIMINARY; PRT; 13 AA.  
 AC Q9TOS1;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)

Thu Jun 6 10:09:20 2002

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91133738; PubMed=2284104;  
 RA Hannink M., Temin H.M.;  
 RT "Structure and autoregulation of the c-rel promoter.";  
 RL Oncogene 5:1843-1850(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Hannink M., Temin H.M.;  
 RL Oncogene 0:0-0(1990).  
 DR EMBL: X56440; CAA39822.1; -;  
 DR EMBL: X56515; CAA39866.1; -;  
 FT NON\_TER 9  
 SQ SEQUENCE 9 AA; 805 MW; DE317DB87865A2CD CRC64;

Query Match 26.6%; Score 17; DB 13; Length 9;  
 Best Local Similarity 42.9%; Pred. No. 5.6e+05;  
 Matches 3; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Qy 2 LVLSSGA 8  
 Db 1 MAVSGGA 7

## RESULT 13

Q98YS3 PRELIMINARY; PRT; 11 AA.  
 AC Q98YS3; (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE TRUNCATED POL PROTEIN (FRAGMENT).  
 GN POL.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=985829;  
 RA Schmidt B., Walter H., Moschik G., Paatz C., Werwein M., Schwingel E.,  
 RA Korn K.;  
 RT "Recovery of HIV-1 pol gene sequences by direct sequencing of  
 RT amplification products derived from plasma samples.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF347394; AAK32471.1; -;  
 FT NON\_TER 1  
 SQ SEQUENCE 11 AA; 1195 MW; E96941B8D878773A CRC64;

Query Match 26.6%; Score 17; DB 15; Length 11;  
 Best Local Similarity 33.3%; Pred. No. 1.2e+04;  
 Matches 3; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 LVLSSGARQ 10  
 Db 3 IXLGGOLRE 11

## RESULT 14

P83054 PRELIMINARY; PRT; 12 AA.  
 AC P83054;  
 DT 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
 DE HEPARIN LYASE (EC 4.2.2.7) (HEPARIN ELIMINASE) (HEPARINASE)  
 DE (FRAGMENT).  
 OS Bacteroides stercoris.

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE TRANSFERRIN (FRAGMENT).  
 OS Equus caballus (Horse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 OX NCBI\_TaxID=9796;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Giffard J.M., Brandon R.B., Bell T.K.;  
 RT "Further identification of single nucleotide polymorphisms in the  
 RT equine transferrin gene.";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF185747; AAF05461.1; -;  
 DR EMBL: AF185746; AAF05460.1; -;  
 FT NON\_TER 1  
 FT NON\_TER 13  
 SQ SEQUENCE 13 AA; 1351 MW; D893DAE970B6CEBD CRC64;

Query Match 28.1%; Score 18; DB 6; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 9e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SGSA 8  
 Db 3 SGSA 6

## RESULT 11

Q9PWP4 PRELIMINARY; PRT; 13 AA.  
 AC Q9PWP4;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE CHIMERIC AFGP/TRYPSTINOGEN-LIKE SERINE PROTEASE (FRAGMENT).  
 OS Disostichus mawsoni (Antarctic cod).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Perciformes; Perciformes;  
 OC Nototheniidae; Nototheniidae; Disostichus.  
 OX NCBI\_TaxID=36200;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99447034; PubMed=10519545;  
 RA Cheng C.H., Chen L.;  
 RT "Evolution of an antifreeze glycoprotein.";  
 RL Nature 401:443-444(1999).  
 DR EMBL: AF134320; AAD37246.1; -;  
 KW Protease.  
 FT NON\_TER 13  
 SQ SEQUENCE 13 AA; 1340 MW; 5186FA54AF1E2727 CRC64;

Query Match 28.1%; Score 18; DB 13; Length 13;  
 Best Local Similarity 57.1%; Pred. No. 9e+03; 1; Indels 0; Gaps 0;  
 Matches 4; Conservative 2; Mismatches 2

Qy 2 LVLSSGA 8  
 Db 7 LLLIGAA 13

## RESULT 12

Q92009 PRELIMINARY; PRT; 9 AA.  
 AC Q92009;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE C-REL PROTEIN (FRAGMENT).  
 OS Gallus gallus (Chicken).

Search completed: June 6, 2002, 06:26:20  
Job time: 193 sec

```

OC Bac'eria; CFB group; Bacteroidetes; Bacteroidales; Bacteroidaceae;
OX NCBI_TaxID=46506;
RN [1]
RP SEQUENCE.
RC STRAIN=HJ-15;
RX MEDLINE=20381043; PubMed=10920269;
RA Kim B.-T., Kim W.-S., Kim Y.S., Linhardt R.J., Kim D.-H.;
RT "Purification and characterization of a novel heparinase from
RT Bacteroides stercoris HJ-15.";
RL J. Biochem. 128:323-328(2000).
CC -|- FUNCTION: DEGRADES HEPARIN AND HEPARAN SULFATE.
CC -|- CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF POLYSACCHARIDES
CC CONTAINING 1,4-LINKED GLUCURONATE OR IDURONATE RESIDUES AND 1,4-
CC ALPHA-LINKED 2-SULFOAMINO-2-DEOXY-6-SULFO-D-GLUCOSE RESIDUES TO
CC GIVE OLIGOSACCHARIDES WITH TERMINAL 4-DEOXY-ALPHA-D-GLUC-4-
CC ENURONOSYL GROUPS AT THEIR NON-REDUCING ENDS.
CC -|- ENZYME REGULATION: INHIBITED BY CUPRIC ION, LEAD AND SOME AGENTS
CC THAT MODIFY HISTIDINE AND CYSTEINE RESIDUES. ACTIVATED BY KCL AND
CC BY REDUCING AGENTS, SUCH AS DITHIOTHREITOL AND 2-MERCAPTOETHANOL.
CC -|- PTM: THE N-TERMINUS IS BLOCKED.
CC -|- MISCELLANEOUS: THE ENZYME HAS AN ISOELECTRIC POINT OF 8.7. ITS
CC OPTIMUM PH IS 7.2 AND THE OPTIMUM TEMPERATURE IS 45 DEGREES
CC CELSIUS.
KW Lyase; Heparin-binding.
FT NON_TER 1
FT NON_TER 12 12
SQ SEQUENCE 12 AA; 1381 MW; CD9CCDB98F6D72D CRC64;

Query Match 26.6%; Score 17; DB 2; Length 12;
Best Local Similarity 40.0%; Pred. No. 1.3e+04;
Matches 4; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 LSGSARQLTF 13
DB 1 MADEALQHTF 10

RESULT 15
Q9TOY4
ID Q9TOY4 PRELIMINARY; PRT; 12 AA.
AC Q9TOY4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE GLYCOPROTEIN H-B N-TERMINAL, GPH-B N-TERMINAL-KEX2/SUBTILISIN-RELATED
DE PROTEASE.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE.
RX MEDLINE=91340701; PubMed=1874725;
RA Christie D.L., Batchelor D.C., Palmer D.J.;
RT "Identification of kex2-related proteases in chromaffin granules by
RT partial amino acid sequence analysis.";
RL J. Biol. Chem. 266:15679-15683(1991).
SQ SEQUENCE 12 AA; 1303 MW; 9F2FF9E2782DC5BA CRC64;

Query Match 26.6%; Score 17; DB 6; Length 12;
Best Local Similarity 55.6%; Pred. No. 1.3e+04;
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 VLGSARQL 11
DB 2 VLXDSALDL 10

```

us-08-881-509-6.closed.rspt

---

Thu Jun 6 10:09:20 2002

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:17:26 ; Search time 29.12 seconds  
(without alignments)  
49.587 Million cell updates/sec

Title: US-08-881-509-6

Perfect score: 64

Sequence: 1 CLVLGSGARQLTF 13

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 166670

Minimum DB seq length: 0

Maximum DB seq length: 13

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	64	100.0	13	19	AAW47590	T-cell receptor CD
2	42.5	66.4	12	19	AAW47591	T-cell receptor CD
3	26.5	41.4	10	22	AAG86220	Saccharomyces cere
4	26.5	41.4	10	22	AAG86310	Saccharomyces cere
5	26	40.6	9	22	AA846917	Human SART-1 deriv
6	26	40.6	10	22	AA844200	Arabidopsis thalia
7	26	40.6	10	22	AAG84202	Arabidopsis thalia
8	26	40.6	11	19	AAW54287	Cytohesin I PH dom
9	25	39.1	9	20	AAV48993	Membrane dipeptida
10	25	39.1	9	22	AA845862	Human tumor-associ
11	25	39.1	10	22	AA856480	Human complementar

12	25	39.1	10	22	AAW96482	Human complementar
13	25	39.1	12	18	AAW32678	Human platelet gly
14	25	39.1	12	19	AAW71808	Mimotope capable o
15	25	39.1	12	22	AA84372	Splice-variance re
16	24	37.5	7	22	AA80722	Human glandular ka
17	24	37.5	8	22	AAW2237	HIV peptide SEQ ID
18	24	37.5	10	22	AAW4861	Human complementar
19	24	37.5	10	22	AAW4863	Human complementar
20	24	37.5	10	22	AAW4863	Human complementar
21	24	37.5	10	22	AAW4863	Human complementar
22	24	37.5	12	15	AAW60702	N-terminal a.a. of
23	24	37.5	12	21	AAW96716	Human E3 ubiquitin
24	23	35.9	6	19	AAW56952	Enzyme inhibitor p
25	23	35.9	6	19	AAW56952	Enzyme inhibitor p
26	23	35.9	9	15	AAW56952	Peptide (155) inh
27	23	35.9	9	15	AAW56952	Peptide (161) inh
28	23	35.9	9	15	AAW56952	Peptide fragment (
29	23	35.9	9	15	AAW56952	Human CASB618 prot
30	23	35.9	9	15	AAW56952	Peptide derived fr
31	23	35.9	9	21	AAW26335	Hepatitis C virus
32	23	35.9	9	22	AAW26335	Hepatitis C virus
33	23	35.9	9	22	AAW26335	Hepatitis C virus
34	23	35.9	9	22	AAW26335	Hepatitis C virus
35	23	35.9	9	22	AAW26335	Hepatitis C virus
36	23	35.9	9	22	AAW26335	Hepatitis C virus
37	23	35.9	10	20	AAW47961	Immunogenic peptid
38	23	35.9	10	21	AAW26383	Human CASB618 prot
39	23	35.9	10	22	AAW42890	Mycoplasma genital
40	23	35.9	10	22	AAW42890	Mycoplasma genital
41	23	35.9	10	22	AAW42890	Human complementar
42	23	35.9	10	22	AAW42890	Human complementar
43	23	35.9	10	22	AAW42890	Saccharomyces cere
44	23	35.9	10	22	AAW42890	Hepatitis C virus
45	23	35.9	10	22	AAW42890	Hepatitis C virus

## ALIGNMENTS

RESULT 1  
AAW47590  
ID AAW47590 standard; peptide; 13 AA.  
AC AAW47590;  
DT 26-JUN-1998 (first entry)  
XX T-cell receptor CDR3 alpha-region.  
XX Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;  
XX Prevention; therapy; tumour disease; renal cell carcinoma;  
XX CDR3.  
OS Homo sapiens.  
XX DE19625191-A1.  
XX 02-JAN-1998.  
XX 24-JUN-1996; 96DE-1025191.  
XX 24-JUN-1996; 96DE-1025191.  
XX (BOEF) BOEHRINGER MANNHEIM GMBH.  
XX Schendel D;  
XX WPI; 1998-053442/06.  
XX N-PSDB; AAV18707.  
XX Human T-cell receptor nucleic acids and poly.peptide(s) - for  
XX diagnosis or therapy, especially of renal cell carcinoma

Thu Jun 6 10:09:14 2002

PS Example 1; Page 17; 30pp; German.

XX The present sequence is the CDR3 alpha-region of a human  
CC T-cell receptor (TCR), which can be used in the diagnosis,  
CC monitoring, prevention and therapy of a tumour disease,  
CC specifically renal cell carcinoma.

XX Sequence 13 AA;

Query Match 100.0%; Score 64; DB 19; Length 13;  
Best Local Similarity 100.0%; Pred. No. 6.2e+06;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CIVLSGSARQLTF 13  
Db 1 civlsgsarqltf 13

RESULT 2  
AAW47591  
ID AAW47591 standard; peptide; 12 AA.

XX AAW47591;

XX 26-JUN-1998 (first entry)  
XX T-cell receptor CDR3 alpha-region.

XX Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;  
XX prevention; therapy; tumour disease; renal cell carcinoma;  
XX CDR3.

XX Homo sapiens.

XX DEL9625191-A1.

XX 02-JAN-1998.

XX 24-JUN-1996; 96DE-1025191.

XX 24-JUN-1996; 96DE-1025191.

XX (BOEF ) BOEHRINGER MANNHEIM GMBH.

XX Schendel D;

XX WPI; 1998-053442/06.

XX N-PSDB; AAV18708.

XX Human T-cell receptor nucleic acids and poly:peptide(s) - for  
XX diagnosis or therapy, especially of renal cell carcinoma

XX Example 1; Page 17; 30pp; German.

XX The present sequence is the CDR3 alpha-region of a human  
CC T-cell receptor (TCR), which can be used in the diagnosis,  
CC monitoring, prevention and therapy of a tumour disease,  
CC specifically renal cell carcinoma.

XX Sequence 12 AA;

Query Match 66.4%; Score 42.5; DB 19; Length 12;  
Best Local Similarity 76.9%; Pred. No. 0.1;  
Matches 10; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 1 CIVLSGSARQLTF 13  
Db 1 cla-tgsarqltf 12

RESULT 3

AAG86220  
ID AAG86220 standard; Peptide; 10 AA.

XX AAG86220;

XX 11-SEP-2001 (first entry)

XX Saccharomyces cerevisiae peptide, SEQ ID NO: 1169.

XX Saccharomyces cerevisiae; complementary peptide; peptide identification;  
XX drug discovery; drug design.

XX Saccharomyces cerevisiae.

XX WO200142276-A1.

XX 14-JUN-2001.

XX 13-DEC-2000; 2000WO-GB04773.

XX 13-DEC-1999; 99GB-0029471.

XX (PROT-) PROTEOM LTD.

XX Roberts GW, Heal JR;

XX WPI; 2001-367863/38.

XX Identifying complementary peptides by analysis of protein and  
XX nucleotide sequence databases, useful in drug design

XX Example 3; Page 191; 488pp; English.

XX The invention relates to the identification of complementary peptides  
CC by analysis of protein and nucleotide sequence databases from higher  
CC eukaryotic genomes, excluding human and plants. The specific  
CC complementary peptides interact with their relevant target proteins  
CC encoded in the eukaryote genome. The peptides may be used as reagents  
CC and drugs for drug discovery and as lead ligands for drug design and  
CC development. The present sequence is a complementary peptide from  
CC Saccharomyces cerevisiae.

XX Sequence 10 AA;

Query Match 41.4%; Score 26.5; DB 22; Length 10;  
Best Local Similarity 70.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 1 CIVL-SGSAR 9  
Db 1 cvlcsgrtar 10

RESULT 4  
AAG86310  
ID AAG86310 standard; Peptide; 10 AA.

XX AAG86310;

XX 11-SEP-2001 (first entry)

XX Saccharomyces cerevisiae peptide, SEQ ID NO: 1259.

XX Saccharomyces cerevisiae; complementary peptide; peptide identification;  
XX drug discovery; drug design.

XX Saccharomyces cerevisiae.

XX WO200142276-A1.

XX 14-JUN-2001.

XX

PF 13-DEC-2000; 2000WO-GB04773.

XX 13-DEC-1999; 99GB-0029471.

XX (PROT-) PROTEOM LTD.

XX Roberts GW, Heal JR;

XX WPI; 2001-367863/38.

XX Identifying complementary peptides by analysis of protein and

XX nucleotide sequence databases, useful in drug design -

XX Example 3; Page 203; 488pp; English.

XX The invention relates to the identification of complementary peptides  
CC by analysis of protein and nucleotide sequence databases from higher  
CC eukaryotic genomes, excluding human and plants. The specific  
CC complementary peptides interact with their relevant target proteins  
CC encoded in the eukaryote genome. The peptides may be used as reagents  
CC and drugs for drug discovery and as lead ligands for drug design and  
CC development. The present sequence is a complementary peptide from  
CC *Saccharomyces cerevisiae*.

XX Sequence 10 AA;

Query Match 41.4%; Score 26.5; DB 22; Length 10;

Best Local Similarity 70.0%; Pred. No. 1.2e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 1 CLVL-SGSAR 9

Db 1 cfvlcsgrar 10

RESULT 5

AAB46917

ID AAB46917 standard; peptide; 9 AA.

XX AAB46917;

XX 04-MAY-2001 (first entry)

XX Human SART-1 derived tumor antigenic peptide SEQ ID 3.

XX Tumor antigenic peptide; interferon; IFN; antigen-specific T cell;  
KW virucide; cytostatic; gene therapy; IFN-alpha; cytotoxic T cell;  
KW viral infectious disease; SART-1; human.

XX Homo sapiens.

XX EP1074267-A1.

XX 07-FEB-2001.

XX 24-JUL-2000; 2000EP-0306263.

XX 22-JUL-1999; 99JP-0207687.

XX (SUMU) SUMITOMO PHARM CO LTD.

XX Takasu H, Gotoh M, Yamaoka T;

XX WPI; 2001-193144/20.

XX Use of antigenic proteins, peptides, interferon or their encoding DNA,  
PT in the manufacture of an agent for the induction of antigen-specific T  
PT cells -

XX Disclosure; Page 12; 25pp; English.

XX This invention describes the novel use of interferons (IFNs) or DNAs

CC capable of expressing the interferons and/or antigenic proteins (AP),  
CC antigenic peptides derived from the proteins or DNAs capable of  
CC expressing the antigenic proteins or peptides, in the manufacture of  
CC an agent for induction of antigen-specific T cells. The products of  
CC the invention have virucide and cytostatic activity and can be used for  
CC gene therapy or as inducers of antigen-specific T cells. The action of  
CC interferon-alpha (IFN-alpha) in a system for inducing specific cytotoxic  
CC T cell (CTL) by administering an antigenic peptide in an incomplete  
CC Freund's adjuvant (IFA) emulsion preparation form was evaluated. IFNs (or  
CC DNA encoding IFNs) are useful in the manufacture of a medicament for  
CC inducing antigen-specific T cells in an individual who has been  
CC administered with AP (or DNA encoding AP) or vice versa. The medicament  
CC is useful for the treatment or prophylaxis of a tumor or a viral  
CC infectious disease.

XX Sequence 9 AA;

Query Match 40.6%; Score 26; DB 22; Length 9;

Best Local Similarity 55.6%; Pred. No. 6.4e+05;

Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 VLGSARQL 11

Db 1 vlsgsgksm 9

RESULT 6

AAG84200

ID AAG84200 standard; Peptide; 10 AA.

XX AAG84200;

XX 11-SEP-2001 (first entry)

XX Arabidopsis thaliana peptide ligand #840.

XX Plant; peptide pesticide; peptide herbicide; agricultural research.

XX Arabidopsis thaliana.

XX WO200142279-A2.

XX 14-JUN-2001.

XX 13-DEC-2000; 2000WO-GB04781.

XX 13-DEC-1999; 99GB-0029469.

XX (PROT-) PROTEOM LTD.

XX Roberts GW, Heal JR;

XX WPI; 2001-381629/40.

XX A set of peptide ligands for agricultural research and development of  
PT therapeutic agents comprise specific complementary peptides to proteins  
PT encoded by genes of plant genomes -

XX Example 4; Page 148; 201pp; English.

XX The present invention relates to a set of peptide ligands consisting of  
CC specific complementary peptides to proteins encoded by genes of plant  
CC genomes. The present sequence is one such peptide from Arabidopsis  
CC thaliana. The peptides of the present invention are useful in an assay to  
CC identify a peptide, especially a peptide pesticide or herbicide. The  
CC peptides are also useful for tools for agricultural research and  
XX development.

XX Sequence 10 AA;

Query Match 40.6%; Score 26; DB 22; Length 10;

Thu Jun 6 10:09:14 2002

us-08-881-509-6.closed.rag

functional molecular surface; protein structural template;  
vaccine; gene therapy; cytohesin 1; human; plasmid pPHCY1;  
random mutagenesis.

functional molecular surface; protein structural template;  
vaccine; gene therapy; cytohesin 1; human; plasmid pPHCY1;  
random mutagenesis.

Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VLSGSA 8  
Db 2 vlsgsa 7

OS Synthetic.  
XX Key Location/Qualifiers  
XX 4..8  
FT Peptide  
FT /label= AB\_loop

RESULT 7  
AAG84202  
ID AAG84202 standard; Peptide: 10 AA.  
XX AC  
XX AAG84202;  
XX 11-SEP-2001 (first entry)  
DE Arabidopsis thaliana peptide ligand #842.  
XX Plant; peptide pesticide; peptide herbicide; agricultural research.  
XX Arabidopsis thaliana.  
OS Arabidopsis thaliana.  
XX WO200142279-A2.  
XX 14-JUN-2001.  
XX 13-DEC-2000; 2000WO-GB04781.  
XX 13-DEC-1999; 99GB-0029469.  
XX (PROT-) PROTEOM LTD.  
XX Roberts GW, Heal JR;  
XX WPI; 2001-381629/40.  
XX A set of peptide ligands for agricultural research and development of  
PT therapeutic agents comprise specific complementary peptides to proteins  
PT encoded by genes of plant genomes -  
XX Example 4; Page 148; 201pp; English.

XX 5 Peptide sequences (see AAW54286-90), respectively designated RM-1,  
RM-18, RM-21, RM-23 and RM-28, represent a peptide library derived  
from the AB loop (see AAW54291) of the pleckstrin homology (PH)  
domain (see AAW54285) of human cytohesin 1. They are encoded by  
5 in-frame mutant sequences (see AAV26482-86) generated by random  
mutagenesis of the DNA region (see AAV26487) encoding the AB loop.  
The average free energy of folding of mutant loop AB peptide RM-18  
is -27 kJ/mol, compared with -39 kJ/mol for the progenitor  
sequence. The average free energy of folding of the 5 mutant  
peptides was -32.4 kJ/mol. Randomisation does not compromise  
the structural integrity or the folding stability of the progenitor  
domain. The invention provides vectors that are used for the  
production of PH domain-like peptide libraries, which can be  
screened to identify peptides that have desirable properties,  
especially novel binding or catalytic properties, and which may be  
of use in research or therapy, or as vaccines. Novel synthetic  
protein structural templates for the generation, screening and  
evolution of functional molecular surfaces are provided.

XX Vectors used to produce PH domain-like peptide libraries - which are  
screened for therapeutically useful peptide(s), e.g. to produce  
vaccines

Example 2; Page 56; 137pp; English.

XX 5 Peptide sequences (see AAW54286-90), respectively designated RM-1,  
RM-18, RM-21, RM-23 and RM-28, represent a peptide library derived  
from the AB loop (see AAW54291) of the pleckstrin homology (PH)  
domain (see AAW54285) of human cytohesin 1. They are encoded by  
5 in-frame mutant sequences (see AAV26482-86) generated by random  
mutagenesis of the DNA region (see AAV26487) encoding the AB loop.  
The average free energy of folding of mutant loop AB peptide RM-18  
is -27 kJ/mol, compared with -39 kJ/mol for the progenitor  
sequence. The average free energy of folding of the 5 mutant  
peptides was -32.4 kJ/mol. Randomisation does not compromise  
the structural integrity or the folding stability of the progenitor  
domain. The invention provides vectors that are used for the  
production of PH domain-like peptide libraries, which can be  
screened to identify peptides that have desirable properties,  
especially novel binding or catalytic properties, and which may be  
of use in research or therapy, or as vaccines. Novel synthetic  
protein structural templates for the generation, screening and  
evolution of functional molecular surfaces are provided.

XX Sequence 11 AA;

XX Sequence 10 AA;

Query Match 40.6%; Score 26; DB 22; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 40.6%; Score 26; DB 19; Length 11;  
Best Local Similarity 62.5%; Pred. No. 1.6e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VLSGSA 8  
Db 3 vlsgsa 8

QY 6 GSARQLTF 13  
Db 3 ghareltw 10

RESULT 8  
AAW54287  
ID AAW54287 standard; Peptide: 11 AA.  
XX AC  
XX AAW54287;  
XX 28-AUG-1998 (first entry)  
DE Cytohesin 1 PH domain randomised AB loop RM-18.  
XX Pleckstrin homology domain; PH domain; peptide library;

RESULT 9  
AAW48993  
ID AAW48993 standard; Peptide: 9 AA.  
XX AC  
XX AAW48993;  
XX 10-DEC-1999 (first entry)

XX Membrane dipeptidase-binding lymph node homing peptide #67.  
XX Homing peptide; organ; tissue; lung; pancreas; skin; retina; MDP;



KW prostate; ovary; lymph node; adrenal gland; liver; gut; tumour;  
 KW membrane dipeptidase.  
 XX Synthetic.  
 OS Homo sapiens.  
 XX WO9946284-A2.  
 XX 16-SEP-1999.  
 XX 10-MAR-1999; 99WO-US05284.  
 XX 13-MAR-1998; 98US-0042107.  
 PR 26-FEB-1999; 99US-0042107.  
 XX (BURN-) BURNHAM INST.  
 XX Rajotte D, Pasqualini R, Ruoslahti EI;  
 XX WPI; 1999-571717/48.  
 XX New peptides which selectively home to organs or tissues, used for,  
 PT e.g. identifying target ligands and for therapy of pathological  
 PT conditions -  
 XX Example 6; Page 154; 193pp; English.  
 XX The present invention describes peptides that selectively home to a  
 CC tissue or organ. The peptides can be used for identifying an organ  
 CC or tissue, for identifying a target molecule expressed by an organ or  
 CC tissue or for treating an organ or tissue pathology, where the organ or  
 CC tissue is selected from prostate, lung, skin, retina, pancreas, gut,  
 CC ovary, adrenal gland, liver, and lymph node. The peptide bind to the  
 CC membrane dipeptidase (MDP). AAY48618 to AAY49066 represent sequences  
 CC which are used in the exemplification of the present invention.  
 XX  
 SQ Sequence 9 AA;  
 Query Match 39.1%; Score 25; DB 20; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CLVLS 5  
 Db 3 CLVLS 7  
 RESULT 10  
 AAB45862  
 ID AAB45862 standard; Protein; 9 AA.  
 AC AAB45862;  
 XX 21-MAR-2001 (first entry)  
 DT Human tumor-associated antigen C42 protein immunogenic fragment #9.  
 DE Tumor-associated antigen; C42; human; immunogenic; cancer therapy;  
 KW cytotatic; immunotherapy; vaccine; lung; breast; esophagus.  
 XX Homo sapiens.  
 OS DE19924199-A1.  
 PN 30-NOV-2000.  
 PD 27-MAY-1999; 99DE-1024199.  
 PF 27-MAY-1999; 99DE-1024199.  
 PR (BOEH ) BOEHRINGER INGELHEIM INT GMBH.  
 XX

PI Adolf G, Heider K, Koenig U, Sommergruber W, Adolf GR, Heider K;  
 PI Koenig U, Sommergruber W;  
 XX WPI; 2001-062549/08.  
 XX Tumor-associated antigen C42, and related nucleic acids and antibodies,  
 PT useful in immunotherapy of cancer and for diagnosis -  
 XX Claim 5; Page 25; 40pp; German.  
 PS This invention describes a novel tumor-associated antigen C42 which has  
 CC cytostatic activity. The invention also describes (a) immunogenic protein  
 CC fragments or peptides (I) derived from C42; (b) a pharmaceutical  
 CC composition for parenteral, topical, oral or local administration  
 CC containing at least 1 C42 or (I); (c) an isolated DNA (II) encoding C42  
 CC or its fragments; (d) a recombinant DNA (Iia) that contains (II); and  
 CC (e) antibodies (Ab) raised against C42 or (I). C42 and its fragments  
 CC induce a humoral immune response and, when presented by major  
 CC histocompatibility complex molecules, a cellular immune response. C42,  
 CC and its fragments, are used for immunotherapy of cancers that express C42  
 CC (particularly of lung, breast and esophagus) to raise specific  
 CC antibodies (Ab) and for diagnosis, e.g. to detect induction of an immune  
 CC response and for optimization of treatment. Ab are used therapeutically  
 CC (e.g. when conjugated to a cytotoxin) or for diagnosis or monitoring of  
 CC cancers that express C42. Nucleic acid (II) that encodes C42 can also be  
 CC used for immunotherapy and cells that express C42 as antitumor vaccines.  
 XX  
 SQ Sequence 9 AA;  
 Query Match 39.1%; Score 25; DB 22; Length 9;  
 Best Local Similarity 55.6%; Pred. No. 6.4e+05;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CLVLSGSAR 9  
 Db 1 CLVLSGSAR 9  
 RESULT 11  
 AAG96480  
 ID AAG96480 standard; Peptide; 10 AA.  
 XX AAG96480;  
 AC AAG96480;  
 XX 18-SEP-2001 (first entry)  
 DT Human complementary peptide, SEQ ID NO: 2674.  
 DE Human; complementary peptide; ligand; drug discovery; drug design.  
 KW Homo sapiens.  
 XX WO200142277-A2.  
 PN 14-JUN-2001.  
 PD 13-DEC-2000; 2000WO-GB04776.  
 PF 13-DEC-1999; 99GB-0029464.  
 PR (PROT-) PROTEOM LTD.  
 XX Roberts GW, Heal JR;  
 PI WPI; 2001-408419/43.  
 XX A set of peptide ligands consisting of specific complementary peptides  
 PT to proteins encoded by genes of the human genome, useful in an assay  
 PT for screening and identifying of one or more novel peptides which are  
 PT drug candidates or pro-drugs -  
 XX Example 4; Page 425; 646pp; English.

Thu Jun 6 10:09:14 2002

XX The invention relates to a set of complementary peptide ligands  
 CC generated from the human genome. The complementary peptides  
 CC interact with their relevant target proteins encoded in the human  
 CC genome. They can be used as reagents in drug discovery and as lead  
 CC ligands to facilitate drug design and development. The present  
 CC sequence is a complementary peptide provided in the specification.  
 XX  
 SQ Sequence 10 AA;

Query Match 39.1%; Score 25; DB 22; Length 10;  
 Best Local Similarity 83.3%; Pred. No. 2.3e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5 SGSARQ 10  
 :|||||  
 Db 4 agsarq 9

RESULT 12  
 AAG96482  
 ID AAG96482 standard; Peptide; 10 AA.  
 XX  
 AC AAG96482;  
 DT 18-SEP-2001 (first entry)  
 XX  
 DE Human complementary peptide, SEQ ID NO: 2676.  
 XX  
 KW Human; complementary peptide; ligand; drug discovery; drug design.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200142277-A2.  
 XX  
 PD 14-JUN-2001.  
 XX  
 PF 13-DEC-2000; 2000WO-GB04776.  
 XX  
 PR 13-DEC-1999; 99GB-0029464.  
 XX  
 PA (PROT-) PROTEOM LTD.  
 XX  
 PI Roberts GW, Heal JR;  
 XX  
 DR WPI; 2001-408419/43.  
 XX  
 PT A set of peptide ligands consisting of specific complementary peptides  
 PT to proteins encoded by genes of the human genome, useful in an assay  
 PT for screening and identifying of one or more novel peptides which are  
 PT drug candidates or pro-drugs -  
 XX  
 PS Example 4; Page 425; 646pp; English.

XX The invention relates to a set of complementary peptide ligands  
 XX generated from the human genome. The complementary peptides  
 CC interact with their relevant target proteins encoded in the human  
 CC genome. They can be used as reagents in drug discovery and as lead  
 CC ligands to facilitate drug design and development. The present  
 CC sequence is a complementary peptide provided in the specification.  
 XX  
 SQ Sequence 10 AA;

Query Match 39.1%; Score 25; DB 22; Length 10;  
 Best Local Similarity 83.3%; Pred. No. 2.3e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5 SGSARQ 10  
 :|||||  
 Db 5 agsarq 10

RESULT 13  
 AAW32678

ID AAW32678 standard; peptide; 12 AA.

XX  
 AC AAW32678;

DT 30-JAN-1998 (first entry)

XX Human platelet glycoprotein Ib/IX complex epitope mimotope.

XX Human platelet glycoprotein Ib/IX complex epitope; mimotope; probe;  
 KW monoclonal antibody C-34; anti mimotope; binding site; antibody;  
 KW platelet adhesion; aggregation; agglutination; von Willebrand factor;  
 KW anti thrombotic drug; ristocetin.  
 XX

OS Synthetic.

OS Homo sapiens.

XX  
 PN WO9718236-A1.

XX  
 PD 22-MAY-1997.

XX  
 PF 08-NOV-1996; 96WO-US17882.

XX  
 PR 13-NOV-1995; 95US-0556597.

XX (UYNV ) UNIV NEW YORK STATE RES FOUND.

XX Lyle VA, Miller JL;

XX WPI; 1997-289227/26.

XX Peptide which mimics human platelet glycoprotein Ib/IX complex  
 PT epitope - also anti-mimotope molecule capable of binding to the  
 PT peptide and useful to modulate platelet adhesion, aggregation or  
 PT agglutination  
 XX  
 PS Claim 3; Page 97; 115pp; English.

XX The present sequence represents a mimotope which functionally mimics a  
 CC binding site for a monoclonal antibody (MAb) which recognises an epitope  
 CC within the human platelet glycoprotein (gp) Ib/IX complex. More  
 CC specifically the MAb is C-34. The mimotope peptide can be used for  
 CC raising antibodies, as probes to search for anti-mimotopes and to  
 CC neutralise the inhibitory activity of those antibodies which recognise  
 CC the epitope which is mimicked by the peptide. Anti-mimotopes can be used  
 CC to modulate the adhesion, aggregation or agglutination of platelets by  
 CC affecting von Willebrand factor interaction with the platelets through  
 CC the gp Ib/IX receptor, e.g. as an anti-thrombotic drug which inhibits  
 CC the ristocetin induced aggregation of platelets.  
 XX  
 SQ Sequence 12 AA;

Query Match 39.1%; Score 25; DB 18; Length 12;  
 Best Local Similarity 66.7%; Pred. No. 2.9e+02;  
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLVLSG 6  
 :|||:  
 Db 4 clvmgs 9

RESULT 14  
 AAW71808  
 ID AAW71808 standard; peptide; 12 AA.

XX  
 AC AAW71808;

DT 08-DEC-1998 (first entry)

XX Mimotope capable of binding to monoclonal antibody C-34 #57.  
 DE

XX Minotope; monoclonal antibody C-34; human; glycoprotein Ib/IX complex;  
 KW binding site; platelet epitope; neutralise; aggregation; inhibition;  
 KW von Willebrand factor interaction; adhesion; agglutination.  
 XX Synthetic.  
 OS Homo sapiens.  
 XX US5817748-A.  
 XX 06-OCT-1998.  
 XX 17-MAR-1995; 95US-0406330.  
 XX 17-MAR-1995; 95US-0406330.  
 XX (UYN ) UNIV NEW YORK STATE RES FOUND.  
 XX Lyle VA, Miller JL;  
 PI WPI; 1998-556458/47.  
 XX Peptides that mimic platelet epitope - and neutralise  
 PT aggregation-inhibiting antibody  
 XX Disclosure; Column 5; 26pp; English.  
 XX The present invention provides peptides which functionally mimic a  
 CC binding site for a monoclonal antibody, where the monoclonal antibody  
 CC recognises an epitope within the human glycoprotein Ib/IX complex.  
 CC Peptides AAW1752 to AAW1832 are mimotopes which are capable of binding  
 CC to monoclonal antibody C-34. The invention also provides an isolated  
 CC molecule capable of binding the mimotopes, where the molecule can be  
 CC an antibody, a second peptide, a carbohydrate, a DNA molecule, an RNA  
 CC molecule, or other naturally or chemically synthesised molecules. These  
 CC molecules are called anti-mimotopes and can act as anti-thrombotic  
 CC drugs. A method has also been described for modulating the adhesion,  
 CC aggregation, or agglutination of platelets, where the method comprises  
 CC selecting platelets and exposing them to an anti-mimotope molecule.  
 CC exposure affects von Willebrand factor interaction with platelets  
 CC through the glycoprotein Ib/IX receptor, which modulates the adhesion,  
 CC aggregation, or agglutination of the platelets. The peptide mimotopes  
 CC mimic platelet epitopes.  
 XX Sequence 12 AA;  
 SQ  
 Query Match 39.1%; Score 25; DB 19; Length 12;  
 Best Local Similarity 66.7%; Pred. No. 2.9e+02;  
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 CLVLSG 6  
 |||: |  
 Db 4 clvmgg 9  
 RESULT 15  
 AAB84372  
 ID AAB84372 standard; peptide; 12 AA.  
 AC AAB84372;  
 XX 22-AUG-2001 (first entry)  
 XX Splice-variance region of a human transferrin.  
 DE Human; transferrin; autoimmune disease; rheumatoid arthritis;  
 XX hyperproliferative disorder; neoplasm; cardiovascular disorder;  
 KW cardiac arrest; cerebrovascular disorder; cerebral ischemia;  
 KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;  
 KW ocular disorder; corneal infection; wound healing;  
 KW epithelial cell proliferation; aging; organ transplant.

OS Homo sapiens.  
 XX WO200146254-A1.  
 PN 28-JUN-2001.  
 XX 21-DEC-2000; 2000WO-US34769.  
 XX 23-DEC-1999; 99US-0171595.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 XX Ruben SM, Shi Y, Choi GH;  
 PI WPI; 2001-381910/40.  
 DR Isolated nucleic acid molecule encoding a human transferrin protein is  
 XX used in preventing, treating or ameliorating a medical condition -  
 PS Disclosure; Page 11; 290pp; English.  
 XX The present sequence is a splice-variance region of a transferrin.  
 CC Transferrin polypeptides and polynucleotides are used to prevent,  
 CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,  
 CC goats, horses, cats, dogs, chickens or sheep. Disorders which are  
 CC diagnosed or treated include autoimmune diseases e.g. rheumatoid  
 CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast  
 CC or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular  
 CC disorders e.g. cerebral ischemia, angiogenesis, nervous system disorders  
 CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and  
 CC fungi and ocular disorders e.g. corneal infection. The polypeptides can  
 CC also be used to aid wound healing and epithelial cell proliferation, to  
 CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities.  
 XX Sequence 12 AA;  
 SQ  
 Query Match 39.1%; Score 25; DB 22; Length 12;  
 Best Local Similarity 50.0%; Pred. No. 2.9e+02;  
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 CLVLSGSA 8  
 | : | | |  
 Db 3 clalkgea 10

Search completed: June 6, 2002, 06:23:03  
 Job time: 337 sec

Thu Jun 6 10:09:14 2002

us-08-881-509-6.closed.rag

---

Page 8

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:17:52 ; Search time 12.85 seconds  
(without alignments)  
24.711 Million cell updates/sec

Title: US-08-881-509-6

Perfect score: 64

Sequence: 1 CLVSGSARLTF 13

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 88308

Minimum DB seq length: 0

Maximum DB seq length: 13

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/6C\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/6D\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	39.1	9	4	US-09-258-754-363
2	25	39.1	9	4	US-09-042-107-363
3	25	39.1	12	2	US-08-406-330-57
4	25	39.1	12	2	US-08-556-597-57
5	24	37.5	12	2	US-08-487-675-2
6	24	37.5	12	2	US-08-904-263A-2
7	23	35.9	13	2	US-08-479-846-4
8	23	35.9	13	2	US-08-454-418A-4
9	23	35.9	13	3	US-08-089-397A-5
10	23	35.9	13	5	PCT-US93-04399-4
11	22	34.4	6	1	US-08-201-046A-4
12	22	34.4	8	1	US-08-526-710-32
13	22	34.4	8	3	US-08-862-855-32
14	22	34.4	8	4	US-09-227-906-32
15	22	34.4	8	4	US-09-226-985-32
16	22	34.4	9	2	US-08-482-651-10
17	22	34.4	9	2	US-08-482-651-25
18	22	34.4	9	4	US-08-660-092-13
19	22	34.4	9	4	US-08-660-092-28
20	22	34.4	9	6	5217869-105
21	22	34.4	12	1	US-07-834-902-6
22	22	34.4	12	1	US-08-018-994-6
23	22	34.4	12	1	US-08-294-675A-6
24	22	34.4	12	3	US-09-188-579-3
25	22	34.4	12	4	US-09-315-444-3
26	21	32.8	8	2	US-08-669-284B-32
27	21	32.8	9	2	US-08-482-651-8

28	21	32.8	9	2	US-08-482-651-35
29	21	32.8	9	4	US-08-660-092-11
30	21	32.8	9	4	US-08-660-092-38
31	21	32.8	9	4	US-08-660-092-59
32	21	32.8	9	4	US-09-510-738A-42
33	21	32.8	9	4	US-09-510-738A-58
34	21	32.8	9	4	US-09-510-738A-99
35	21	32.8	10	1	US-08-250-789A-4
36	21	32.8	10	1	US-08-346-333-61
37	21	32.8	10	2	US-08-968-542C-20
38	21	32.8	10	3	US-08-925-002-7
39	21	32.8	10	4	US-09-306-593-3
40	21	32.8	10	4	US-09-230-548-3
41	21	32.8	10	5	PCT-US91-07506-61
42	21	32.8	11	2	US-08-748-021-22
43	21	32.8	11	2	US-08-771-602D-36
44	21	32.8	11	3	US-08-904-446A-9
45	21	32.8	11	3	US-08-974-297-22

## ALIGNMENTS

## RESULT 1

US-09-258-754-363  
; Sequence 363, Application US/09258754  
; Patent No. 6174687

; GENERAL INFORMATION:

; APPLICANT: Ruoslahti, Erkki

; APPLICANT: Pasqualini, Renata

; APPLICANT: Rajotte, Daniel

; TITLE OF INVENTION: Methods of Identifying Lung Homing Molecules Using

; FILE REFERENCE: P-LJ 3443

; CURRENT APPLICATION NUMBER: US/09/258,754

; CURRENT FILING DATE: 1999-02-26

; EARLIER APPLICATION NUMBER: 09/042,107

; EARLIER FILING DATE: 1998-03-13

; NUMBER OF SEQ ID NOS: 452

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 363

; LENGTH: 9

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Synthetic

US-09-258-754-363

## Query Match

Best Local Similarity 39.1%; Score 25; DB 4; Length 9;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY 1 CLVLS 5

Db 3 CLVLS 7

## RESULT 2

US-09-042-107-363

; Sequence 363, Application US/09042107

; Patent No. 6232287

; GENERAL INFORMATION:

; APPLICANT: Ruoslahti, Erkki

; APPLICANT: Pasqualini, Renata

; TITLE OF INVENTION: Molecules that Home to Various Selected Organs or

; FILE REFERENCE: P-LJ 2892

; CURRENT APPLICATION NUMBER: US/09/042,107

; CURRENT FILING DATE: 1998-03-13

; NUMBER OF SEQ ID NOS: 436

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 363

```

;
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; US-09-042-107-363

Query Match          39.1%; Score 25; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVLS 5
Db 3 CLVLS 7

RESULT 3
US-08-406-330-57
; Sequence 57, Application US/08406330
; Patent No. 5817748
; GENERAL INFORMATION:
; APPLICANT: Miller, Jonathan L.
; APPLICANT: Lyle, Vicki A.
; TITLE OF INVENTION: MIMOTOPES AND ANTI-MIMOTOPES OF
; HUMAN PLATELET GLYCOPROTEIN Ib/IX
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/556,597
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/406,330
; FILING DATE: 17-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Timian, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20884/101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1636
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-556-597-57

Query Match          39.1%; Score 25; DB 2; Length 12;
Best Local Similarity 66.7%; Pred. No. 72;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLVLSG 6
Db 4 CLVMG 9

RESULT 5
US-08-487-675-2
; Sequence 2, Application US/08487675
; Patent No. 5869064
; GENERAL INFORMATION:
; APPLICANT: LINDAHL, Gunnar
; APPLICANT: STALHAMMAR-CARLEMALM, Margaretha
; APPLICANT: STENBERG, Lars
; TITLE OF INVENTION: PROTEIN RIB, A CELL SURFACE PROTEIN
; THAT CONFERS IMMUNITY TO MANY STRAINS OF THE GROUP
; B STREPTOCOCCUS: PROCESS FOR PURIFICATION OF THE
; TITLE OF INVENTION: B STREPTOCOCCUS: REAGENT KIT AND PHARMACEUTICAL COMPOSITION
; TITLE OF INVENTION: PROTEIN, REAGENT KIT AND PHARMACEUTICAL COMPOSITION
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

;
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; US-09-042-107-363

Query Match          39.1%; Score 25; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLVLS 5
Db 3 CLVLS 7

RESULT 3
US-08-406-330-57
; Sequence 57, Application US/08406330
; Patent No. 5817748
; GENERAL INFORMATION:
; APPLICANT: Miller, Jonathan L.
; APPLICANT: Lyle, Vicki A.
; TITLE OF INVENTION: MIMOTOPES AND ANTI-MIMOTOPES OF
; HUMAN PLATELET GLYCOPROTEIN Ib/IX
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/406,330
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20884/100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1636
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-406-330-57

Query Match          39.1%; Score 25; DB 2; Length 12;
Best Local Similarity 66.7%; Pred. No. 72;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLVLSG 6
Db 4 CLVMG 9

RESULT 4
US-08-556-597-57
; Sequence 57, Application US/08556597
; Patent No. 5877155

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,675
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/SE94/00246
; FILING DATE: 21-MAR-1994
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/SE93/00234
; FILING DATE: 19-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 552-114P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus group B
; STRAIN: BM110
;
US-08-487-675-2

```

```

Query Match      37.5%  Score 24;  DB 2;  Length 12;
Best Local Similarity 83.3%  Pred. No. 1.1e+02;
Matches 5;  Conservative 1;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

QY 3 VLSGSA 8
   1:|||||
DB 3 VISGSA 8

```

```

RESULT 6
US-08-904-263A-2
; Sequence 2, Application US/08904263A
; Patent No. 6015889
;
; GENERAL INFORMATION:
; APPLICANT: LINDAHL, GUNNAR
; APPLICANT: STALHAMMAR-CARLEWALM,
; APPLICANT: MARGARETHA
; APPLICANT: STENBERG, LARS
; TITLE OF INVENTION: PROTEIN RTB, A CELL SURFACE PROTEIN THAT
; TITLE OF INVENTION: CONFERS IMMUNITY TO MANY STRAINS OF THE GROUP B
; TITLE OF INVENTION: STREPTOCOCCUS: PROCESS FOR PURIFICATION OF THE PROTEIN,
; TITLE OF INVENTION: REAGENT KIT AND PHARMACEUTICAL COMPOSITION
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/904,263A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: SVENSSON, LEONARD R
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 552-119P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..921
;
US-08-904-263A-2

```

```

Query Match      37.5%  Score 24;  DB 3;  Length 12;
Best Local Similarity 83.3%  Pred. No. 1.1e+02;
Matches 5;  Conservative 1;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

QY 3 VLSGSA 8
   1:|||||
DB 3 VISGSA 8

```

```

RESULT 7
US-08-479-846-4
; Sequence 4, Application US/08479846
; Patent No. 5874534
;
; GENERAL INFORMATION:
; APPLICANT: Vegeto, Elisabetta
; APPLICANT: McDonnell, Donald P.
; APPLICANT: O'Malley, Bert W.
; APPLICANT: Schrader, William T.
; APPLICANT: Tsai, Ming-Jer
; TITLE OF INVENTION: Mutated Steroid Hormone Receptors,
; TITLE OF INVENTION: Methods for Their Use and Molecular Switch for Gene
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77010-3095
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,846
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/939,246
; FILING DATE: 02-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D
; REGISTRATION NUMBER: 32,714
; REFERENCE/DOCKET NUMBER: D-5505
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/651-5151
; TELEFAX: 713/651-5246
; TELEX: 762829

```

us-08-881-509-6.closed.ra1

Thu Jun 6 10:09:14 2002

```

; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: US 07/882,771
; FILING DATE: 14-MAY-1992
; US-08-479-846-4

Query Match 35.9%; Score 23; DB 2; Length 13;
Best Local Similarity 42.9%; Pred. No. 2e+02;
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLVLSGS 7
Db 2 CMILSNN 8

RESULT 8
US-08-454-418A-4
; Sequence 4, Application US/08454418A
; Patent No. 5935934
; GENERAL INFORMATION:
; APPLICANT: Vegeto, Elisabetta
; APPLICANT: McDonnell, Donald P.
; APPLICANT: O'Malley, Bert W.
; APPLICANT: Schrader, William T.
; APPLICANT: Tsai, Ming-Jer
; TITLE OF INVENTION: MUTATED STEROID HORMONE RECEPTORS,
; TITLE OF INVENTION: METHODS FOR THEIR USE AND MOLECULAR
; TITLE OF INVENTION: SWITCH FOR GENE THERAPY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,418A
; FILING DATE: May 30, 1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/939,246
; FILING DATE: September 2, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkman, Charles S.
; REGISTRATION NUMBER: 38,077
; REFERENCE/DOCKET NUMBER: 212/295
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

; MOLECULE TYPE: Peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: US 07/882,771
; FILING DATE: May 14, 1992
; US-08-454-418A-4

Query Match 35.9%; Score 23; DB 2; Length 13;
Best Local Similarity 42.9%; Pred. No. 2e+02;
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLVLSGS 7
Db 2 CMILSNN 8

RESULT 9
US-08-089-397A-5
; Sequence 5, Application US/08089397A
; Patent No. 6086880
; GENERAL INFORMATION:
; APPLICANT: SABARA, MARTA I.J.
; APPLICANT: FRENCHICK, PATRICK J.
; APPLICANT: POTTER, ANDREW A.
; APPLICANT: IJAZ, MOHAMMAD K.
; APPLICANT: GILCHRIST, JAMES E.
; APPLICANT: REDMOND, MARK J.
; TITLE OF INVENTION: ROTAVIRUS VACCINES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/089,397A
; FILING DATE: 07-JUL-1993
; CLASSIFICATION: 330
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 29311-20003.03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-089-397A-5

Query Match 35.9%; Score 23; DB 3; Length 13;
Best Local Similarity 37.5%; Pred. No. 2e+02;
Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 6 GSARQLTF 13
Db 2 GASQIIV 9
```



RESULT 10  
PCT-US93-04399-4  
; Sequence 4, Application PC/TUS9304399  
; GENERAL INFORMATION:  
; APPLICANT: Vegeto, Elisabetta  
; APPLICANT: McDonnell, Donald P.  
; APPLICANT: O'Malley, Bert W.  
; APPLICANT: Schrader, William T.  
; APPLICANT: Tsai, Ming-Jer  
; TITLE OF INVENTION: Mutated Steroid Hormone Receptors,  
; TITLE OF INVENTION: Methods for Their Use and Molecular Switch for Gene  
; TITLE OF INVENTION: Therapy  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fulbright & Jaworski  
; STREET: 1301 McKinney, Suite 5100  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: U.S.A.  
; ZIP: 77010-3095  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/04399  
; FILING DATE: 19930511  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/939,246  
; FILING DATE: 02-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul, Thomas D.  
; REGISTRATION NUMBER: 32,714  
; REFERENCE/DOCKET NUMBER: D-5505  
; TELEPHONE: 713/651-5151  
; TELEFAX: 713/651-5246  
; TELEX: 762829  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 13 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; PUBLICATION INFORMATION:  
; DOCUMENT NUMBER: US 07/882,771  
; FILING DATE: 14-MAY-1992  
PCT-US93-04399-4

Query Match 35.9%; Score 23; DB 5; Length 13;  
Best Local Similarity 42.9%; Pred. No. 26+02;  
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLVLGSG 7  
|:|:|:  
Db 2 CMILSNN 8

RESULT 11  
US-08-201-046A-4  
; Sequence 4, Application US/08201046A  
; Patent No. 5545719  
; GENERAL INFORMATION:  
; APPLICANT: Shashoua, Victor E.  
; TITLE OF INVENTION: NERVE GROWTH PEPTIDES AND USES THEREFOR  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Ave.  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/201,046A  
; FILING DATE: 24-FEB-1994  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: N0260/7013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: Internal  
; US-08-201-046A-4

Query Match 34.4%; Score 22; DB 1; Length 6;  
Best Local Similarity 66.7%; Pred. No. 1.7e+05;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLVLGSG 6  
|:|:|:  
Db 1 CLPVSG 6

RESULT 12  
US-08-526-710-32  
; Sequence 32, Application US/08526710  
; Patent No. 5622699  
; GENERAL INFORMATION:  
; APPLICANT: Ruoslahti, Erkki  
; APPLICANT: Pasqualini, Renata  
; TITLE OF INVENTION: Method of Identifying Molecules That  
; TITLE OF INVENTION: Home to a Selected Organ In Vivo  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/526,710  
; FILING DATE: 11-SEP-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 1779  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001

Db 1 CKLLSG 6

RESULT 14

US-09-226-985-32

Sequence 32, Application US/09226985

Patent No. 6296832

GENERAL INFORMATION:

APPLICANT: Ruoslahti, Erkki

APPLICANT: Pasqualini, Renata

TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo

NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell & Flores LLP

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: United States

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/226,985

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/526,710

FILING DATE: 11-SEP-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/813,273

FILING DATE: 10-MAR-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/862,855

FILING DATE: 23-MAY-1997

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-LJ 3423

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 8 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-226-985-32

Query Match 34.4%; Score 22; DB 4; Length 8;

Best Local Similarity 66.7%; Pred. No. 1.7e+05;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CKLLSG 6

Db 1 CKLLSG 6

RESULT 15

US-09-227-906-32

Sequence 32, Application US/09227906

Patent No. 6306365

GENERAL INFORMATION:

APPLICANT: Ruoslahti, Erkki

APPLICANT: Pasqualini, Renata

TITLE OF INVENTION: Method of Identifying Molecules That

TITLE OF INVENTION: Home to a Selected Organ In Vivo

NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESS:

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 8 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-526-710-32

Query Match 34.4%; Score 22; DB 1; Length 8;

Best Local Similarity 66.7%; Pred. No. 1.7e+05;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CKLLSG 6

Db 1 CKLLSG 6

RESULT 13

US-08-862-855-32

Sequence 32, Application US/08862855

Patent No. 6068829

GENERAL INFORMATION:

APPLICANT: Ruoslahti, Erkki

APPLICANT: Pasqualini, Renata

TITLE OF INVENTION: Method of Identifying Molecules That

TITLE OF INVENTION: Home to a Selected Organ In Vivo

NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell & Flores LLP

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: United States

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/862,855

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/526,710

FILING DATE: 11-SEP-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/813,273

FILING DATE: 10-MAR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-LJ 2621

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 8 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-862-855-32

Query Match 34.4%; Score 22; DB 3; Length 8;

Best Local Similarity 66.7%; Pred. No. 1.7e+05;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CKLLSG 6

```

: ADDRESSEE: Campbell & Flores LLP
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: United States
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/227,906
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/526,710
: FILING DATE: 11-SEP-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/813,273
: FILING DATE: 10-MAR-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/862,855
: FILING DATE: 23-MAY-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LJ 3424
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 32:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 8 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-09-227-906-32

```

```

Query Match      34.4%; Score 22; DB 4; Length 8;
Best Local Similarity 66.7%; Pred. No. 1.7e+05;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

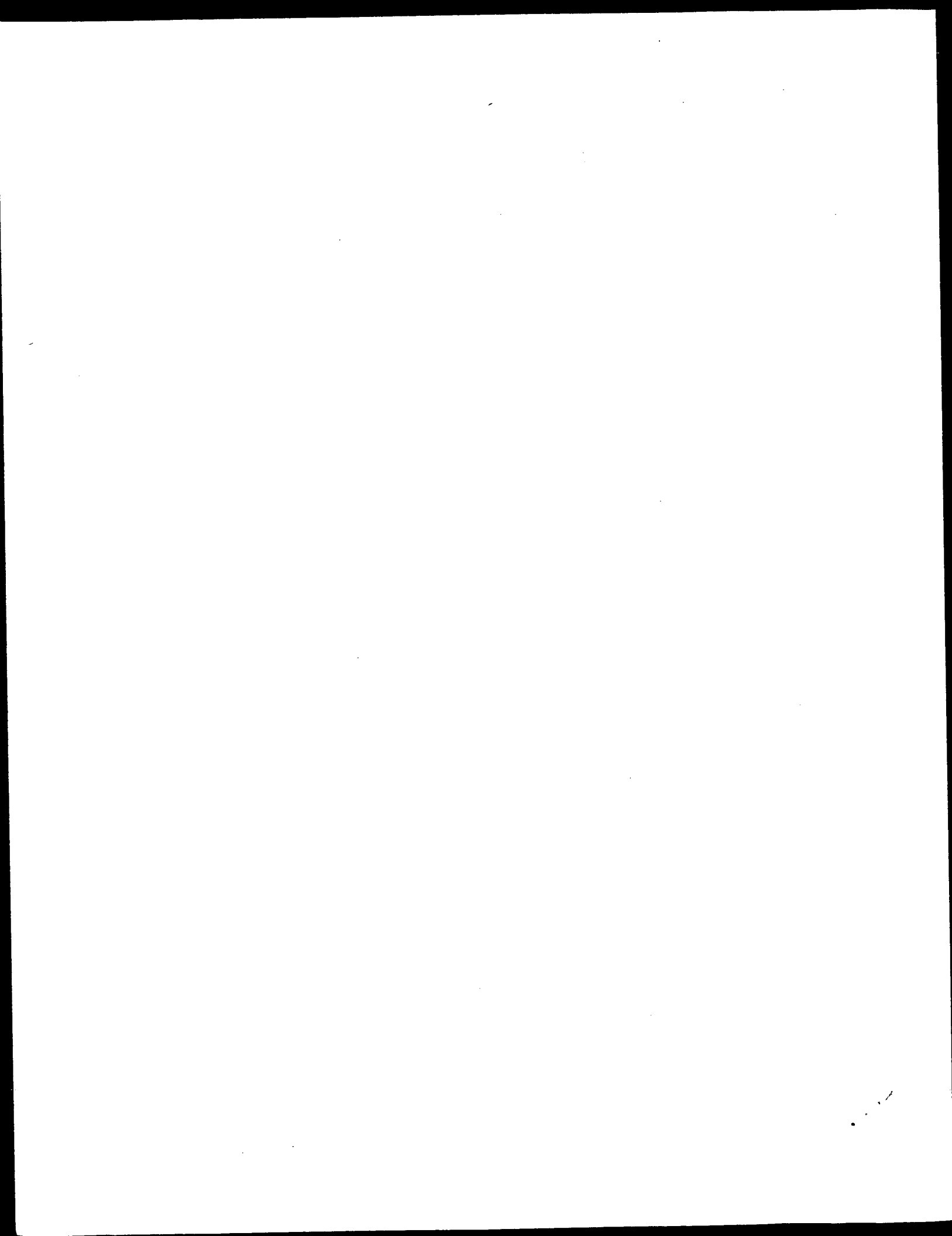
QY 1 CLVLSG 6
   | :|||
Db 1 CKLLSG 6

```

```

Search completed: June 6, 2002, 06:23:22
Job time: 330 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 07:39:13 ; Search time 2149.71 seconds  
(without alignments)  
350.445 Million cell updates/sec

Title: US-08-881-509-7

Perfect score: 36

Sequence: 1 TGCCCTGCTACTGGTCTGCAAGCAACTGACCTTT 36

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.inv.\*

32: em.htg.other.\*

33: em.htgo.inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----					

1	36	100.0	36	6	A93133	A93133 Sequence 7
2	36	100.0	1318	9	HSTCRJUNC	X98410 H.sapiens m
3	29.6	82.2	1341	6	A93127	A93127 Sequence 1
4	28.6	79.4	716	9	S63879	S63879 TCR V alpha
5	28	77.8	39	9	HSU30448	U30448 Human Isola
6	28	77.8	51	9	AF043886	AF043886 Homo sapi
7	27.6	76.7	39	6	A93131	A93131 Sequence 5
8	27.4	76.1	39	9	HSU30428	U30428 Human Isola
9	27.4	76.1	51	9	AF043875	AF043875 Homo sapi
10	27.4	76.1	51	9	AF043877	AF043877 Homo sapi
11	27.4	76.1	51	9	AF043884	AF043884 Homo sapi
12	27.4	76.1	63	9	HSU30390	U30390 Human Isola
13	27.4	76.1	411	9	HUMTCVJCC	M97706 Human T-cel
14	27.4	76.1	645	9	HSTCRJ3	X02886 Human gene
15	27.4	76.1	71153	9	HUAE000662	AE000662 Homo sapi
16	27.4	76.1	97630	9	HUMTCRADCV	M94081 Human Tcr-C
17	27.4	76.1	175053	2	AC023226	AC023226 Homo sapi
18	27	75.0	48	9	HSU27254	U27254 Human Isola
19	26.8	74.4	87	9	HUMTCRAGG	L42801 Homo sapien
20	26.4	73.3	42	9	HSU235208	AJ235208 Homo sapi
21	26.4	73.3	51	9	AF043888	AF043888 Homo sapi
22	26.4	73.3	51	9	AF043890	AF043890 Homo sapi
23	26.4	73.3	183	9	HUMTCATA	M17666 Human T-cel
24	26.4	73.3	222	9	AF333590	AF333590 Homo sapi
25	26	72.2	51	9	AF043874	AF043874 Homo sapi
26	26	72.2	51	9	AF043876	AF043876 Homo sapi
27	26	72.2	51	9	AF043892	AF043892 Homo sapi
28	26	72.2	51	9	AF043894	AF043894 Homo sapi
29	26	72.2	98	9	S69140	S69140 Homo sapien
30	25.4	70.6	214	9	HUMIGTCACA	M81774 Homo sapien
31	25	69.4	234	9	AF041171	U40776 Human T-cel
32	24.2	67.2	42	9	HSTCELL14	X81549 H.sapiens r
33	24.2	67.2	51	9	HSU14029	U14029 Human T cel
34	24.2	67.2	51	9	HSTCARAVB	Z22966 H.sapiens m
35	24.2	67.2	246	9	AC046129	AC046129 Homo sapi
36	23.2	64.4	140531	9	AC016101	AC016101 Homo sapi
37	23.2	64.4	196629	2	AC087242	AC087242 Homo sapi
38	23.2	64.4	333458	2	OSJN00048	AL606615 Oryza sat
39	22.2	61.7	165253	2	AC023442	AC023442 Homo sapi
40	21.8	60.6	177686	2	AC020710	AC020710 Homo sapi
41	21.8	60.6	181925	2	AF021956	AF021956 Homo sapi
42	21.8	60.6	197864	9	AF295529	AF295529 Mus muscu
43	21.6	60.0	9157	10	AC109364	AC109364 Mus muscu
44	21.6	60.0	30933	2	AC101555	AC101555 Mus muscu
45	21.6	60.0	64695	2		

## ALIGNMENTS

RESULT 1

A93133

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

linear PAT 22-JAN-2000

A93133 Sequence 7 from Patent EP0816496.

36 bp

GI:6741522

unidentified.

unclassified.

1 (bases 1 to 36)

Schandel,D.J.

T-cells specific for kidney carcinoma

Patent: EP 0816496-A 7 07-JAN-1998;

BOEHRINGER MANNHEIM GMBH (DE)

Location/Qualifiers

1..36

/organism="unidentified"

/db\_xref="taxon:32644"

<1..>36

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAB69532.1"

```

/db_xref="GI:6741523"
/translation="CLATGSARQLTF"
6 a 11 c 8 g 11 t

Query Match 100.0%; Score 36; DB 6; Length 36;
Best Local Similarity 100.0%; Pred. No. 2,2e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TGCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36
|||||
1 TGCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36

RESULT 2
HSTCRJUNC 1318 bp mRNA linear PRI 08-JAN-1997
H:scapiens mRNA for rearranged TCR junctional sequences.
X98410
X98410.1 GI:1770560
J gene; junction; T cell receptor; TCR junctional sequence; V gene.
human.
Organism Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1318)
Jantzer,P.U. and Schendel,D.J.
Tumor-infiltrating lymphocytes recognizing spontaneously arising
renal cell carcinomas express T cell receptors characteristic of a
secondary immune response
Unpublished
2 (bases 1 to 1318)
Jantzer,P.
Direct Submission
Submitted (11-JUN-1996) P. Jantzer, Institute for Immunology,
Goethestr. 31, D- 80336 Munich, FRG
Location/Qualifiers
1. .1318
/organism="Homo sapiens"
/isolate="patients 22 and 26"
/db_xref="taxon:9606"
/rearranged
<1. >1318
/note="V gene/J gene junction"
/codon_start=1
/product="TCR junctional sequence"
/protein_id="CAA67057.1"
/db_xref="GI:1770561"
/translation="CLVGG SARQLTFGXC LATSARQLTFGXC LVLGSGARQLTFGXC
LVLGSGARQLTFGXC LVLGSGARQLTFGXC LVLGSGARQLTFGXC LVLGSGARQLTFGXC
LVLGSGARQLTFGXC LVLGSGARQLTFGXC LVLGSGARQLTFGXC LVLGSGARQLTFGXC
VSGARQLTFGXC LVLGSGARQLTFGXC LVLGSGARQLTFGXC LVLGSGARQLTFGXC
LVAGGNTPLVFGXC LVLGSGARQLTFGXC LVLGSGARQLTFGXC LVLGSGARQLTFGXC
ATGSARQLTFGXC LVLGAGGYQKVTFGXC LVLGSGARQLTFGXC LVLGSGARQLTFGXC
CLVVLGSGQNLFFGXC LVLGPNAGNMLTFGXC LVLATNOGRNCSDLW"

BASE COUNT 206 a 329 g 344 t 87 others
ORIGIN

Query Match 100.0%; Score 36; DB 9; Length 1318;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 36
|||||
DB 43 TGCTCGCTACTGGTTCTGCAAGGCAACTGACCTTT 78

RESULT 3
A93127

```



ORGANISM	HOMO sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 51)
AUTHORS	Striebach,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
TITLE	Selective accumulation of related CD4+ T cell clones in the



```

JOURNAL synovial fluid of patients with rheumatoid arthritis
MEDLINE J. Immunol. 161 (8), 4428-4436 (1998)
REFERENCE 2 (bases 1 to 51)
AUTHORS Striebach,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and
Research Center, 1400 Jackson St., Denver, CO 80206, USA
FEATURES
source
1. .51
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="AV4"
/tissue_type="synovial fluid"
/note="from rheumatoid arthritis patient CS-1 [TCRAV1S3]
[TCRAJ22]"
<1. .>51
/gene="TCRA"
<1. .>51
/note="Contains the 3' end of the TCRAV, the n, and the 5'
end of the TCRAJ."
/codon_start=1
/product="T cell receptor alpha chain CDR3"
/protein_id="AAC72677.1"
/db_xref="GI:3859384"
/translation="YFCAVRPSGSARQLTFG"
BASE COUNT 9 a 12 c 13 g 17 t
ORIGIN

Query Match 76.1%; Score 27.4; DB 9; Length 51;
Best Local Similarity 96.6%; Pred. No. 0.18;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 8 CTACTGGTTCTGCAAGGCAACTGACCTTT 36
|| |||||
Db 20 CTCTGGTTCTGCAAGGCAACTGACCTTT 48

RESULT 11
AF043884 51 bp mRNA linear PRI 11-NOV-1998
LOCUS CDR3 (TCRA) mRNA, partial cds.
DEFINITION Homo sapiens patient CS-1 clone AV325 T cell receptor alpha chain
ACCESSION AF043884
VERSION AF043884.1 GI:3859387
KEYWORDS
SOURCE human.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 51)
AUTHORS Striebach,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
TITLE Selective accumulation of related CD4+ T cell clones in the
synovial fluid of patients with rheumatoid arthritis
JOURNAL J. Immunol. 161 (8), 4428-4436 (1998)
MEDLINE 98451502
REFERENCE 2 (bases 1 to 51)
AUTHORS Striebach,C.C., Falta,M.T., Wang,Y., Bill,J. and Kotzin,B.L.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1998) Pediatrics, National Jewish Medical and
Research Center, 1400 Jackson St., Denver, CO 80206, USA
FEATURES
source
1. .51
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="AV325"
/tissue_type="synovial fluid"
/note="from rheumatoid arthritis patient CS-1 [TCRAV1S2]
[TCRAJ22]"
<1. .>51
/gene="TCRA"

```

```

CDS
<1. .>51
/gene="TCRA"
/note="Contains the 3' end of the TCRAV/TCRBV, the nDn/n,
and the 5' end of the TCRAJ/TCRBJ."
/codon_start=1
/product="T cell receptor alpha chain CDR3"
/protein_id="AAC72679.1"
/db_xref="GI:3859388"
/translation="YFCAVRSGSGSARQLTFG"
BASE COUNT 7 a 12 c 14 g 18 t
ORIGIN

Query Match 76.1%; Score 27.4; DB 9; Length 51;
Best Local Similarity 96.6%; Pred. No. 0.18;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 8 CTACTGGTTCTGCAAGGCAACTGACCTTT 36
|| |||||
Db 20 CTCTGGTTCTGCAAGGCAACTGACCTTT 48

RESULT 12
HSU30390 63 bp mRNA linear PRI 01-AUG-1995
LOCUS Human isolate M101 T-cell receptor alpha V-J junction (TCR Valpha
DEFINITION 5/J alpha 22) mRNA, partial cds.
ACCESSION U30390
VERSION U30390.1 GI:915415
KEYWORDS
SOURCE human.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 63)
AUTHORS Dave,V.P., Larche,M., Rencher,S.D., Koop,B.F. and Hurwitz,J.L.
TITLE Restricted usage of T-cell receptor V alpha sequence and
variable-joining pairs after normal T-cell development and bone
marrow transplantation
JOURNAL Hum. Immunol. 37 (3), 178-184 (1993)
MEDLINE 94064390
REFERENCE 2 (bases 1 to 63)
AUTHORS Hurwitz,J.L.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-1995) Julia L. Hurwitz, St. Jude Children's
Research Hospital, Immunology, 332 N. Lauderdale, Memphis, TN
38101, USA
FEATURES
source
1. .63
Location/Qualifiers
/organism="Homo sapiens"
/isolate="M101"
/db_xref="taxon:9606"
/tissue_type="blood"
1. .63
/gene="TCR Valpha 5/J alpha 22"
<1. .>63
/gene="TCR Valpha 5/J alpha 22"
/codon_start=1
/product="T-cell receptor alpha V-J junction"
/protein_id="AAA73579.1"
/db_xref="GI:915416"
/translation="TYLCALDPRVSSGSARQLTF"
BASE COUNT 11 a 20 c 12 g 20 t
ORIGIN

Query Match 76.1%; Score 27.4; DB 9; Length 63;
Best Local Similarity 96.6%; Pred. No. 0.18;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 8 CTACTGGTTCTGCAAGGCAACTGACCTTT 36
|| |||||
Db 35 CTCTGGTTCTGCAAGGCAACTGACCTTT 63

```

```

RESULT 13
HUMTCVJCC 411 bp mRNA linear PRI 14-JAN-1995
LOCUS Human T-cell receptor (V alpha 22.1, J alpha C, C alpha 1) mRNA.
DEFINITION M97706
ACCESSION M97706.1 GI:292803
VERSION T-cell receptor; T-cell receptor alpha.
KEYWORDS Homo sapiens (individual_isolate AL61.270) cDNA to mRNA.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Hurley,C.K., Steiner,N., Wagner,A., Geiger,M.J., Eckels,D.D. and
Rosen-Bronson,S.
TITLE Nonrandom T cell receptor usage in the allorecognition of HLA-DR1
microvariation
J. Immunol. 150 (4), 1314-1324 (1993)
JOURNAL 93163556
MEDLINE *haplotype = DRB1*0101
COMMENT Reference [J. Immunol. (1993) In press] reports bases 324 to 393.
FEATURES
Source Location/Qualifiers
1..411
/organism="Homo sapiens"
/isolate="AL61.270"
/db_xref="taxon:9606"
/map="14q11.2"
/haplotype="DRB1*0101"
/cell_type="alloproliferative T lymphocyte clone"
<1..53
/sig_peptide
/gene="TCRA"
/note="G00-120-404; does not fit consensus"
/product="T-cell receptor alpha"
1..411
/gene="TCRA"
54..323
/gene="TCRA"
/standard_name="V alpha 22.1"
/note="G00-120-404; putative"
/citation=[1]
/product="T-cell receptor alpha"
324..393
/gene="TCRA"
/standard_name="J alpha C"
/note="Includes CDR3 region and J segment; G00-120-404; putative"
/citation=[1]
/product="T-cell receptor alpha"
394..411
/gene="TCRA"
/standard_name="Constant alpha 1"
/note="G00-120-404; putative"
/citation=[1]
/product="T-cell receptor alpha"
BASE COUNT 107 a 106 c 92 g 106 t
ORIGIN
Query Match 76.1%; Score 27.4; DB 9; Length 411;
Best Local Similarity 96.6%; Pred. No. 0.16;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0
QY 8 CTACTGGTCTGCAAGGCAACTGACCTTT 36
|| |||||
DB 334 CTTCCTGGTCTGCAAGGCAACTGACCTTT 362
|| |||||
RESULT 14
HSTCRAJ3 HSTCRAJ3 645 bp DNA linear PRI 11-APR-1995
LOCUS Human gene for T-cell receptor alpha chain J region.
DEFINITION X02886
ACCESSION X02886

```

Search completed: June 6, 2002, 07:39:15  
Job time: 5378 sec

Thu Jun 6 10:09:26 2002

us-08-881-509-7.rge

---

Page 8

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 06:54:29 ; Search time 2649.09 seconds  
(without alignments)  
183.418 Million cell updates/sec

Title: US-08-881-509-7  
Perfect score: 36  
Sequence: 1 TCCTCGCTACTGTTCTGCAAGCACTGACCTTT 36

Scoring table: IDENTITY\_NUC  
Gap 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26.4	73.3	564	12 AQ221133	AQ221133 HS_3223_B
2	21.8	60.6	1060	12 CNS04W88	AL310049 Tetraodon
3	21.6	60.0	284	9 AI715777	AI715777 UI-R-YO-a
4	21.2	58.9	282	9 BB491342	BB491342 BB491342
5	21.2	58.9	316	12 AZ492378	AZ492378 IM0326B02
6	21.2	58.9	1002	12 CNS079XE	AL435864 T7 end of
7	21.2	58.9	1088	12 CNS077IC	AL432118 T7 end of
8	20.8	57.8	393	9 AA019771	AA019771 ze62e03.r
9	20.8	57.8	488	12 AQ423489	AQ423489 CITHI-EI-
10	20.8	57.8	506	12 AQ470385	AQ470385 CITHI-EI-
11	20.8	57.8	667	10 BF304050	BF304050 601886829
12	20.8	57.8	852	10 BG479738	BG479738 602526929
13	20.6	57.2	440	12 AZ232465	AZ232465 4908.gf21
14	20.6	57.2	527	10 BI534791	BI534791 397951.MA
15	20.6	57.2	667	9 BB639586	BB639586 BB639586
16	20.6	57.2	874	10 BF242338	BF242338 601880546
17	20.4	56.7	120	10 BI423478	BI423478 949048D12

C 18	20.4	56.7	371	10 BF899667	BF899667 IL5-MT020
C 19	20.4	56.7	547	12 AQ049405	AQ049405 CLM-14c12
C 20	20.4	56.7	619	12 BH062049	BH062049 RPCI-24-3
C 21	20.4	56.7	1098	10 BI912745	BI912745 G03176136
C 22	20.2	56.1	462	12 BH236190	BH236190 AT2KB19TF
C 23	20.2	56.1	475	12 BH235936	BH235936 AT2KE49TR
C 24	20.2	56.1	586	12 BH236089	BH236089 AT2KE92TR
C 25	20.2	56.1	618	9 BB615995	BB615995 BB615995
C 26	20.2	56.1	657	12 BH235920	BH235920 AT2KB30TF
C 27	20.2	56.1	664	12 BH235999	BH235999 AT2KE32TF
C 28	20.2	56.1	694	12 BH235850	BH235850 AT2KF36TR
C 29	20.2	56.1	939	10 BG366394	BG366394 HVSMEL000
C 30	20.2	56.1	960	9 AL581298	AL581298 AL581298
C 31	20.2	56.1	985	12 CNS04DT3	AL286176 Tetraodon
C 32	20.2	56.1	1015	10 BE744953	BE744953 G01576221
C 33	20	55.6	219	10 BM191333	BM191333 dae17004.
C 34	20	55.6	283	9 AV171596	AV171596 AV171596
C 35	20	55.6	288	9 AV147538	AV147538 AV147538
C 36	20	55.6	301	9 AV136764	AV136764 AV136764
C 37	20	55.6	330	10 BI314592	BI314592 dae14f12.
C 38	20	55.6	348	9 AV110678	AV110678 AV110678
C 39	20	55.6	357	10 T97080	T97080 ye50g08.r1
C 40	20	55.6	359	10 BG513863	BG513863 dae03f08.
C 41	20	55.6	402	12 BH074362	BH074362 RPCI-24-2
C 42	20	55.6	415	10 BG264799	BG264799 daa32g10.
C 43	20	55.6	466	10 BG022611	BG022611 dab06f12.
C 44	20	55.6	477	12 AQ432798	AQ432798 HS_5132_B
C 45	20	55.6	501	10 BG022723	BG022723 dab07h06.

## ALIGNMENTS

## RESULT 1

AQ221133 564 bp DNA linear GSS 19-SEP-1998  
LOCUS HS\_3223\_B2\_E10\_T7 C17 Approved Human Genomic Sperm Library D Homo  
DEFINITION sapiens genomic clone plate=3223 Col=20 Row=J, DNA sequence.  
ACCESSION AQ221133  
VERSION AQ221133.1 GI:3634746  
KEYWORDS GSS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 564)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 3223 row: J column: 20  
Class: BAC ends  
High quality sequence stop: 564.  
Location/Qualifiers  
1. 564  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="plate=3223 Col=20 Row=J"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in  
E-Coli DH10B"

BASE COUNT	160 a	102 c	124 g	171 t	7 others
ORIGIN					
Query Match	73.3%	Score 26.4;	DB 12;	Length 564;	
Best Local Similarity	96.4%;	Pred. No. 2.4;			
Matches	27;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	9	TACTGTGTTCTCAAGGCAACTGACCTTT	36		
DB	423	TTCGTGTTCTCGAAGCAACTGACCTTT	450		
RESULT	2				
CNS04W88/c					
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					









Thu Jun 6 10:09:30 2002

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLCMI001 row: j column: 08  
 High quality sequence stop: 645.

# FEATURES

source

1. .667

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4120807"

/clone\_lib="NIH\_MGC\_17"

/tissue\_type="rhabdomyosarcoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: muscle; Vector: pOTB7; Site\_1: EcoRI; Site\_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 173 a 169 c 174 g 151 t

ORIGIN

Query Match 57.8%; Score 20.8; DB 10; Length 667;

Best Local Similarity 78.1%; Pred. No. 4.3e+02;

Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 5 TCGCTACTGTTCTGCAAGGCAACTGACCTT 36

Db 591 TCCCAACTGTTATGTAAGCAACTTCCCTCT 622

RESULT 12

LOCUS BG479738/c

DEFINITION 602526929F1 NIH\_MGC\_21 Homo sapiens cDNA clone IMAGE:4650702 5', mRNA sequence.

ACCESSION BG479738

VERSION BG479738.1

KEYWORDS GI:13412017

SOURCE EST. human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 852)

AUTHORS NTH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLCMI432 row: i column: 07

High quality sequence stop: 732.

FEATURES

Location/Qualifiers

1. .852

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4650702"

/clone\_lib="NIH\_MGC\_21"

/tissue\_type="choriocarcinoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: placenta; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 173 a 169 c 174 g 151 t

ORIGIN

Query Match 57.8%; Score 20.8; DB 12; Length 506;

Best Local Similarity 78.1%; Pred. No. 3.9e+02;

Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CTCGCTACTGTTCTGCAAGGCAACTGACCTT 35

Db 251 CTCCTACTGTTTCCACTAGGCCACTGAACTT 220

RESULT 11

LOCUS BF304050

DEFINITION 601886829F2 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4120807 5', mRNA sequence.

ACCESSION BF304050

VERSION BF304050.1

KEYWORDS GI:11250727

SOURCE EST. human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 667)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLCMI432 row: i column: 07

High quality sequence stop: 732.

FEATURES

Location/Qualifiers

1. .506

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="2584F3"

/clone\_lib="CITBI-El"

/sex="male"

/cell\_type="sperm"

/note="Vector: pBeloBAC11; Site\_1: EcoRI; Site\_2: EcoRI; Caltech Human BAC Library D"

BASE COUNT 160 a 91 c 127 g 128 t

ORIGIN

Query Match 57.8%; Score 20.8; DB 12; Length 506;

Best Local Similarity 78.1%; Pred. No. 3.9e+02;

Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CTCGCTACTGTTCTGCAAGGCAACTGACCTT 35

Db 251 CTCCTACTGTTTCCACTAGGCCACTGAACTT 220

RESULT 11

LOCUS BF304050

DEFINITION 601886829F2 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4120807 5', mRNA sequence.

ACCESSION BF304050

VERSION BF304050.1

KEYWORDS GI:11250727

SOURCE EST. human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 667)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLCMI432 row: i column: 07

High quality sequence stop: 732.

FEATURES

Location/Qualifiers

1. .506

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="2584F3"

/clone\_lib="CITBI-El"

/sex="male"

/cell\_type="sperm"

/note="Vector: pBeloBAC11; Site\_1: EcoRI; Site\_2: EcoRI; Caltech Human BAC Library D"

BASE COUNT 160 a 91 c 127 g 128 t

ORIGIN

Query Match 57.8%; Score 20.8; DB 12; Length 506;

Best Local Similarity 78.1%; Pred. No. 3.9e+02;

Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CTCGCTACTGTTCTGCAAGGCAACTGACCTT 35

Db 251 CTCCTACTGTTTCCACTAGGCCACTGAACTT 220

RESULT 11

LOCUS BF304050

DEFINITION 601886829F2 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4120807 5', mRNA sequence.

ACCESSION BF304050

VERSION BF304050.1

KEYWORDS GI:11250727

SOURCE EST. human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 667)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLCMI432 row: i column: 07

High quality sequence stop: 732.

FEATURES

Location/Qualifiers

1. .506

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="2584F3"

/clone\_lib="CITBI-El"

/sex="male"

/cell\_type="sperm"

/note="Vector: pBeloBAC11; Site\_1: EcoRI; Site\_2: EcoRI; Caltech Human BAC Library D"

BASE COUNT 160 a 91 c 127 g 128 t

ORIGIN

Query Match 57.8%; Score 20.8; DB 12; Length 506;

Best Local Similarity 78.1%; Pred. No. 3.9e+02;

Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CTCGCTACTGTTCTGCAAGGCAACTGACCTT 35

Db 251 CTCCTACTGTTTCCACTAGGCCACTGAACTT 220

RESULT 11

LOCUS BF304050

DEFINITION 601886829F2 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4120807 5', mRNA sequence.

ACCESSION BF304050

VERSION BF304050.1

KEYWORDS GI:11250727

SOURCE EST. human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 667)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLCMI432 row: i column: 07

High quality sequence stop: 732.

FEATURES

Location/Qualifiers

1. .506

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="2584F3"

/clone\_lib="CITBI-El"

/sex="male"

/cell\_type="sperm"

/note="Vector: pBeloBAC11; Site\_1: EcoRI; Site\_2: EcoRI; Caltech Human BAC Library D"

BASE COUNT 160 a 91 c 127 g 128 t

ORIGIN

Query Match 57.8%; Score 20.8; DB 12; Length 506;

Best Local Similarity 78.1%; Pred. No. 3.9e+02;

Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CTCGCTACTGTTCTGCAAGGCAACTGACCTT 35

Db 251 CTCCTACTGTTTCCACTAGGCCACTGAACTT 220

RESULT 11

LOCUS BF304050

DEFINITION 601886829F2 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4120807 5', mRNA sequence.

ACCESSION BF304050

VERSION BF304050.1

KEYWORDS GI:11250727

SOURCE EST. human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 667)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLCMI432 row: i column: 07

High quality sequence stop: 732.

FEATURES

Location/Qualifiers

1. .506

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="2584F3"

/clone\_lib="CITBI-El"

/sex="male"

/cell\_type="sperm"

/note="Vector: pBeloBAC11; Site\_1: EcoRI; Site\_2: EcoRI; Caltech Human BAC Library D"

BASE COUNT 160 a 91 c 127

of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

```

BASE COUNT      219 a 185 c 177 g 271 t
ORIGIN

Query Match      57.8%; Score 20.8; DB 10; Length 852;
Best Local Similarity 78.1%; Pred. No. 4.6e+02;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 GCCTCGTACTGGTTCTGCAAGCAACTGACC 33
||||| ||||| || ||| |||||
Db 477 GCCTAGTACTGGACCTCAAGACACTGACC 446

RESULT 13
AZ923465/c
LOCUS
DEFINITION
4908.gf21c06.s1 Saccharomyces cariocanus UFRJ 50791 Saccharomyces
cariocanus genomic clone 4908.gf21c06.s1, DNA sequence.
ACCESSION
AZ923465
VERSION
AZ923465.1 GI:13494363
KEYWORDS
GSS.
SOURCE
Saccharomyces cariocanus.
ORGANISM
Saccharomyces cariocanus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE
AUTHORS
Clifton, P.F., Hillier, L.W., Fulton, L., Graves, T., Miner, T., Gish
,W.R., Waterston, R.H. and Johnston, M.
TITLE
Surveying Saccharomyces genomes to identify functional elements by
comparative DNA sequence analysis
JOURNAL
Unpublished (2001)
COMMENT
Contact: Johnston M
Department of Genetics
Washington University Medical School
Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
Tel: 314 362 2735
Fax: 314 362 7855
Email: mj@genetics.wustl.edu
Class: random plasmid subclone.
Location/Qualifiers
1. 440
/organism="Saccharomyces cariocanus"
/strain="UFRJ 50791"
/db_xref="taxon:114526"
/clone="4908.gf21c06.s1"
/clone_lib="Saccharomyces cariocanus UFRJ 50791"
/notes="Random genomic sequence"

BASE COUNT      132 a 69 c 86 g 153 t
ORIGIN

Query Match      57.2%; Score 20.6; DB 12; Length 440;
Best Local Similarity 74.3%; Pred. No. 4.5e+02;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 TGCCTCGTACTGGTTCTGCAAGCAACTGACCTT 35
||||| ||||| || ||||| ||||| |||||
Db 136 TGCACGCTATTGACTCTGTAAGAACTGACCGT 102

RESULT 14
BI534791
LOCUS
DEFINITION
397951.MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION
BI534791
VERSION
BI534791.1 GI:15375899
KEYWORDS
EST.
SOURCE
COW.
ORGANISM
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
```

```

Bovidae; Bovinae; Bos.
1 (bases 1 to 527)
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett
,G.L., Heaton, M.P., Laegreid, W., Rohrer, G.A., Chitko-McKown, C.G.,
Perteau, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keele, J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
COMMENT
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCACGACG
Plate: 123 row: G column: 14
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1. 527
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 4BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="vector; PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 20 and day 40
embryos."

BASE COUNT      132 a 126 c 123 g 146 t
ORIGIN

Query Match      57.2%; Score 20.6; DB 10; Length 527;
Best Local Similarity 74.3%; Pred. No. 4.8e+02;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 TGCCTCGTACTGGTTCTGCAAGCAACTGACCTT 35
||||| ||||| || ||||| ||||| |||||
Db 73 TGCCTCTATAATGCTCTGCAAGACAGCTGTTT 107

RESULT 15
BB639586/c
LOCUS
DEFINITION
BB639586 RIKEN full-length enriched, 7 days neonate cerebellum Mus
musculus cDNA clone A730061C10 5', mRNA sequence.
ACCESSION
BB639586
VERSION
BB639586.1 GI:16474972
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 667)
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,
Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda
,M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Kouda
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki
,D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toyota, T.,
Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
```

The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Fax: 01-43-303-9220  
 Email: genome-res@gsc.riken.go.jp,  
 URL: <http://genome.gsc.riken.go.jp/>  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi, K., Fujiwara, S., Inoue, K., Todawa, Y., Izawa, M., Ohara, E.,  
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura  
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
 Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10  
 (11), 1757-1771 (2000)  
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
 Y. and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamana, I., Aizawa  
 K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and  
 Hayashizaki, Y.  
 Computational Analysis of Full-Length Mouse cDNAs Compared with  
 Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
 please visit our web site (<http://genome.gsc.riken.go.jp>) for  
 further details.

FEATURES  
source

```

Tissues. Location/Qualifiers
1. .667
   /organism="Mus musculus"
   /db_xref="taxon:10090"
   /clone="A730061C10"
   /clone_lib="RIKEN full-length enriched, 7 days neonate
   cerebellum"
   /tissue_type="cerebellum"
   /dev_stage="7 days neonate"
   /lab_host="DH10B"
   /note="Site.1: SalI; Site.2: BamHI; cDNA library was
   prepared and sequenced in Mouse Genome Encyclopedia
   Project of Genome Exploration Research Group in Riken
   Genomic Sciences Center and Genome Science Laboratory in
   RIKEN. Division of Experimental Animal Research in Riken
   contributed to prepare mouse tissues. 1st strand cDNA was
   primed with a primer [5'
   GAGAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
   prepared by using trihalose thermo-activated reverse
   transcriptase and subsequently enriched for full-length by
   cap-trapper. cDNA went through one round of normalization
   to Rot = 20.0 and subtraction to Rot = 459.0. Second
   strand cDNA was prepared with the primer adapter of
   sequence [5' GAGAGAGAGATCTCGAGTTAAATTAATATCCCCCCCCCCCC
   3'], cDNA was cleaved with XhoI and BamHI. Vector: a
   modified pBluescript KS(+) after bulk excision from Lambda
   FLC I."
1 180 c 140 g 145 t 1 others

```

BASE COUNT	201 a	180 c	140 g	145 t	1 others
ORIGIN					
Query Match		57.2%;	Score 20.6;	DB 9;	Length 667;
Best Local Similarity		85.2%;	Pred. No. 5.2e+02;		
Matches	23;	Conservative	0;	Mismatches 4;	Indels 0; Gaps 0;
Dy	5	TCGCTACTGTGTTCTTGC	AAAGCGCAACTGA	31	
DB	124	TCATTAGTGTGTTCTTGC	AAAGCGCAACTGA	98	

search completed: June 6, 2002, 06:54:33

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 6, 2002, 07:00:23 ; Search time 333.99 Seconds  
(without alignments)  
185.062 Million cell updates/sec

Title: US-08-881-509-7

Perfect score: 36

Sequence: 1 TGCTCCTACTGTTCTGCAAGCAACTGACCTT 36

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	100.0	36	AAV18708	CDNA for T-cell re
2	29.6	82.2	1341	AAV18705	CDNA for T-cell re
3	27.6	76.7	39	AAV18707	CDNA for T-cell re
4	19.6	54.4	5482	ABL23904	Drosophila melanog
5	19.4	53.9	133	ABAA47504	Human breast cell
6	19.4	53.9	133	ABA65394	Human breast cell
7	19.4	53.9	133	ABA32492	Human foetal liver
8	19.4	53.9	133	AAK13803	Probe #10958 for g
9	19.4	53.9	133	AAK39548	Human brain expres

10	19.4	53.9	133	AAI20356	Probe #10289 for g
11	19.4	53.9	133	AAI45561	Probe #14247 used t
12	19.4	53.9	133	AAI06058	Probe #6049 used t
13	19.4	53.9	264	ABAI7256	Human nervous syst
14	19.4	53.9	265	ABAI7254	Human nervous syst
15	19.4	53.9	265	ABAI7255	Human nervous syst
16	19.4	53.9	456	ABA42363	Human breast cell
17	19.4	53.9	456	ABA52789	Human foetal liver
18	19.4	53.9	456	ABA22574	Probe #1040 for ge
19	19.4	53.9	456	AAK01043	Human brain expres
20	19.4	53.9	456	AAK26502	Human bone marrow
21	19.4	53.9	456	AAI11130	Probe #1063 for ge
22	19.4	53.9	456	AAI32395	Probe #1081 used t
23	19.4	53.9	456	AAI01048	Probe #1039 used t
24	19	52.8	1001	AAH51555	Human FMO related
25	19	52.8	1473	AAH51555	Pseudomonas sp ABC
26	19	52.8	1632	AAH51555	CDNA sequence #115
27	19	52.8	26016	AAH51555	Human flavin-conta
28	18.8	52.2	1456	AAH51555	LHP polypeptide en
29	18.6	51.7	302	AAH51555	Mycobacterium tube
30	18.6	51.7	303	AAH51555	Mycobacterium tube
31	18.6	51.7	327	AAH51555	Mycobacterium tube
32	18.6	51.7	327	AAH51555	Mycobacterium tube
33	18.6	51.7	327	AAH51555	M. tuberculosis im
34	18.6	51.7	327	AAH51555	Mycobacterium tube
35	18.6	51.7	327	AAH51555	M. tuberculosis an
36	18.6	51.7	327	AAH51555	M. tuberculosis re
37	18.6	51.7	396	AAH51555	Mycobacterium tube
38	18.6	51.7	396	AAH51555	Mycobacterium tube
39	18.6	51.7	396	AAH51555	M. tuberculosis im
40	18.6	51.7	396	AAH51555	Mycobacterium tube
41	18.6	51.7	396	AAH51555	M. tuberculosis an
42	18.6	51.7	396	AAH51555	M. tuberculosis re
43	18.6	51.7	460	AAH51555	Human breast cell
44	18.6	51.7	460	AAH51555	Human foetal liver
45	18.6	51.7	460	AAH51555	Probe #1245 for ge

# ALIGNMENTS

RESULT 1	
AAV18708	
ID	AAV18708 standard; cdNA; 36 BP.
XX	AAV18708;
XX	AC
XX	AC
DT	26-JUN-1998 (first entry)
XX	cdNA for T-cell receptor CDR3 alpha-region.
XX	Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;
KW	prevention; therapy; tumour disease; renal cell carcinoma;
KW	CDR3; ds.
XX	Homo sapiens.
XX	OS
XX	XX
FT	Key
FT	mat_peptide
FT	Location/Qualifiers
XX	1..36
XX	/*tag= a
PN	DE19625191-A1.
XX	XX
PD	02-JAN-1998.
XX	XX
PF	24-JUN-1996; 96DE-1025191.
XX	XX
PR	24-JUN-1996; 96DE-1025191.
XX	XX
PA	(BOEF ) BOEHRINGER MANNHEIM GMBH.
XX	XX
PI	Schandel D;
XX	XX

us-08-881-509-7.rng

Thu Jun 6 10:09:27 2002

```
DR WPI: 1998-053442/06.
DR P-PSDB; AAW47591.
XX Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT diagnosis or therapy, especially of renal cell carcinoma
XX
XX Example 1; Page 17; 30pp; German.
XX The present sequence encodes the CDR3 alpha-region of a human
CC T-cell receptor (TCR), which can be used in the diagnosis,
CC monitoring, prevention and therapy of a tumour disease,
CC specifically renal cell carcinoma.
XX
XX Sequence 36 BP; 6 A; 11 C; 8 G; 11 T; 0 other;
SQ
Query Match 100.0%; Score 36; DB 19; Length 36;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCTCGCTACTGGTTCGCAAGGCAACTGACCTTT 36
| | | | | | | | | | | | | | | | | | | | | |
Db 1 tgctcgtactgggttctgcaaggcaactgaccttt 36
| | | | | | | | | | | | | | | | | | | | | |
RESULT 2
AAV18705
ID AAV18705 standard; cDNA; 1341 BP.
XX
AC AAV18705;
XX
DT 26-JUN-1998 (first entry)
XX
DE cDNA for T-cell receptor alpha-chain.
XX
KW Alpha-chain; human; T-cell receptor; TCR; diagnosis; monitoring;
KW prevention; therapy; tumour disease; renal cell carcinoma; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT mat_peptide 1..54
FT FT /*tag= a
XX
XX DE19625191-A1.
XX
XX 02-JAN-1998.
XX
XX 24-JUN-1996; 96DE-1025191.
XX
XX 24-JUN-1996; 96DE-1025191.
XX
XX (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
XX Schendel D;
XX
XX WPI: 1998-053442/06.
XX P-PSDB; AAW47591.
XX
XX Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT diagnosis or therapy, especially of renal cell carcinoma
XX
XX Example 1; Pages 11-13; 30pp; German.
XX The present sequence encodes the alpha-chain of a human T-cell
CC receptor (TCR), which can be used in the diagnosis, monitoring,
CC prevention and therapy of a tumour disease, specifically renal
CC cell carcinoma.
XX
XX Sequence 36 BP; 6 A; 11 C; 8 G; 11 T; 0 other;
SQ
Query Match 100.0%; Score 36; DB 19; Length 36;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCTCGCTACTGGTTCGCAAGGCAACTGACCTTT 36
| | | | | | | | | | | | | | | | | | | | | |
Db 1 tgctcgtactgggttctgcaaggcaactgaccttt 36
| | | | | | | | | | | | | | | | | | | | | |
RESULT 3
AAV18707
ID AAV18707 standard; cDNA; 39 BP.
XX
AC AAV18707;
XX
DT 26-JUN-1998 (first entry)
XX
DE cDNA for T-cell receptor CDR3 alpha-region.
XX
KW Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;
KW prevention; therapy; tumour disease; renal cell carcinoma;
KW CDR3; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT mat_peptide 1..39
FT FT /*tag= a
XX
XX DE19625191-A1.
XX
XX 02-JAN-1998.
XX
XX 24-JUN-1996; 96DE-1025191.
XX
XX 24-JUN-1996; 96DE-1025191.
XX
XX (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
XX Schendel D;
XX
XX WPI: 1998-053442/06.
XX P-PSDB; AAW47590.
XX
XX Human T-cell receptor nucleic acids and poly:peptide(s) - for
PT diagnosis or therapy, especially of renal cell carcinoma
XX
XX Example 1; Page 17; 30pp; German.
XX
XX The present sequence encodes the CDR3 alpha-region of a human
CC T-cell receptor (TCR), which can be used in the diagnosis,
CC monitoring, prevention and therapy of a tumour disease,
CC specifically renal cell carcinoma.
XX
XX Sequence 39 BP; 5 A; 12 C; 8 G; 14 T; 0 other;
SQ
Query Match 76.7%; Score 27.6; DB 19; Length 39;
Best Local Similarity 88.2%; Pred. No. 0.013;
Matches 30; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 3 CCTCGCTACTGGTTCGCAAGGCAACTGACCTTT 36
| | | | | | | | | | | | | | | | | | | | | |
Db 6 cgtccttctggttctgcaaggcaactgaccttt 39
| | | | | | | | | | | | | | | | | | | | | |
RESULT 4
ABL23904/c
ID ABL23904 standard; DNA; 5482 BP.
```

XX AC ABL23904;  
 XX DT 26-MAR-2002 (first entry)  
 XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 23185.  
 XX KW Drosophila; developmental biology; cell signalling; insecticide;  
 XX KW pharmaceutical; gene; ds.  
 XX OS Drosophila melanogaster.  
 XX PN WO200171042-A2.  
 XX PD 27-SEP-2001.  
 XX PF 23-MAR-2001; 2001WO-US09231.  
 XX PR 23-MAR-2000; 2000US-191637P.  
 XX PR 11-JUL-2000; 2000US-0614150.  
 XX PA (PEKE ) PE CORP NY.  
 XX PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX DR WPI; 2001-656860/75.  
 XX CC New isolated nucleic acid detection reagent for detecting 1000 or more  
 CC genes from Drosophila and for elucidating cell signalling and cell-cell  
 CC interactions.  
 CC Claim 1; SEQ ID NO 23185; 21pp + Sequence Listing; English.  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 5482 BP; 1352 A; 1226 C; 1338 G; 1566 T; 0 other;

Query Match 54.4%; Score 19.6; DB 23; Length 5482;  
 Best Local Similarity 73.5%; Pred. No. 1.1e+02;  
 Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 2 GCCTCGTACTGGTTCTGCAAGGCAACTGACCTT 35  
 ||| | ||||| ||| ||||| |||  
 Db 37 GCCCTCGACTGGTCTGCTGAATGCAAGTGCTTT 4

RESULT 5  
 ABA47504  
 ID ABA47504 standard; DNA; 133 BP.  
 XX AC ABA47504;  
 XX DT 01-FEB-2002 (first entry)  
 XX DE Human breast cell single exon nucleic acid probe #6199.  
 XX KW Human; microarray; single exon probe; gene expression; breast;  
 XX KW disease; cancer; ss.  
 XX OS Homo sapiens.  
 XX PN WO200157271-A2.

XX 09-AUG-2001.  
 XX PD 30-JAN-2001; 2001WO-US00662.  
 XX PF 04-FEB-2000; 2000US-0180312.  
 XX PR 26-MAY-2000; 2000US-0207456.  
 XX PR 30-JUN-2000; 2000US-0608408.  
 XX PR 03-AUG-2000; 2000US-0632366.  
 XX PR 21-SEP-2000; 2000US-0234687.  
 XX PR 27-SEP-2000; 2000US-0236359.  
 XX PR 04-OCT-2000; 2000GB-0024263.  
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX DR WPI; 2001-496933/54.  
 XX PT New spatially-addressable set of single exon nucleic acid probes,  
 XX PT useful for measuring gene expression in sample derived from human  
 XX PT breast, comprises number of single exon nucleic acid probes -  
 XX PS Claim 4; SEQ ID NO 6199; 327pp + sequence listing; English.  
 XX CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human breast and BT 474 cells. The method involves contacting  
 CC the probes with a collection of detectably labelled nucleic acids  
 CC derived from mRNA of human breast, and then measuring the label  
 CC bound to each probe of the microarray. The probes are useful for  
 CC verifying the expression of regions of genomic DNA predicted to  
 CC encode proteins. They are useful for gene discovery, and for  
 CC determining predisposition and/or prognosing breast disease. Gené  
 CC expression analysis is useful for assessing the toxicity of chemical  
 CC agents on cells. The microarray of this invention presents a far greater  
 CC diversity of probes for measuring gene expression, with far less bias  
 CC than expressed sequence tag microarrays. The method is suitable for  
 CC rapid production of functional information from genomic sequence. The  
 CC present sequence is a single exon nucleic acid probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 133 BP; 33 A; 44 C; 26 G; 30 T; 0 other;

Query Match 53.9%; Score 19.4; DB 22; Length 133;  
 Best Local Similarity 79.3%; Pred. No. 63;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 7 GCTACTGTTCTGCAAGGCAACTGACCTT 35  
 || |||| ||||| ||||| |||||  
 Db 46 gccactgtctcagtggaactgtcctt 74

RESULT 6  
 ABA65394  
 ID ABA65394 standard; DNA; 133 BP.  
 XX AC ABA65394;  
 XX DT 01-FEB-2002 (first entry)  
 XX DE Human foetal liver single exon nucleic acid probe #13699.  
 XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
 XX OS Homo sapiens.  
 XX PN WO200157277-A2.  
 XX PD 09-AUG-2001.





PT Single exon nucleic acid probes for analyzing gene expression in human  
 XX brains -

PS Example 4; SEQ ID NO: 13794; 650pp + Sequence Listing; English.

CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is one of the probes of the  
 CC invention.

XX Sequence 133 BP; 33 A; 44 C; 26 G; 30 T; 0 other;

Query Match 53.9%; Score 19.4; DB 22; Length 133;  
 Best Local Similarity 79.3%; Pred. No. 63;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 7 GCTACTGGTTCTGCAAGGCAACTGACCTT 35  
 || |||| |||| |||| |||| |||| ||||  
 Db 46 gccactgcttcgactggaactgtcctt 74

RESULT 9

AAK39548  
 ID AAK39548 standard; DNA; 133 BP.

AC AAK39548;

DT 06-NOV-2001 (first entry)

DE Human bone marrow expressed single exon probe SEQ ID NO: 14105.

KW Human; bone marrow expressed exon; gene expression analysis; probe;  
 microarray; cancer; leukaemia; lymphoma; myeloma; ss.

OS Homo sapiens.

PN WO200157276-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00668.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488900/53.

PT Human genome-derived single exon nucleic acid probes useful for  
 analyzing gene expression in human bone marrow -

PS Example 4; SEQ ID NO: 14105; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
 CC the probes of the invention.

XX Sequence 133 BP; 33 A; 44 C; 26 G; 30 T; 0 other;

Query Match 53.9%; Score 19.4; DB 22; Length 133;  
 Best Local Similarity 79.3%; Pred. No. 63;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 7 GCTACTGGTTCTGCAAGGCAACTGACCTT 35  
 || |||| |||| |||| |||| |||| ||||  
 Db 46 gccactgcttcgactggaactgtcctt 74

RESULT 10

AAI20356  
 ID AAI20356 standard; DNA; 133 BP.

XX AAI20356;

DT 12-OCT-2001 (first entry)

DE Probe #10289 for gene expression analysis in human cervical cell sample.  
 KW Probe; human; microarray; gene expression; cervical epithelial cell;  
 cervical cancer; ss.

OS Homo sapiens.

PN WO200157278-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488901/53.

PT Human genome-derived single exon nucleic acid probes useful for  
 analyzing gene expression in human cervical epithelial cells -

PS Claim 25; SEQ ID No 10289; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes  
 CC (SENPs). The present sequence is one such probe. The SENPs are derived  
 CC from human HeLa cells. The SENPs can be used to produce a single exon  
 CC microarray, which can be used for measuring human gene expression in a  
 CC sample derived from human cervical epithelial cells. By measuring gene  
 CC expression, the probes are therefore useful in grading and/or staging  
 CC of diseases of the cervix, notably cervical cancer.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 133 BP; 33 A; 44 C; 26 G; 30 T; 0 other;

Query Match 53.9%; Score 19.4; DB 22; Length 133;  
 Best Local Similarity 79.3%; Pred. No. 63;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 7 GCTACTGGTTCTGCAAGGCAACTGACCTT 35  
 || |||| |||| |||| |||| |||| ||||  
 Db 46 gccactgcttcgactggaactgtcctt 74

Thu Jun 6 10:09:27 2002

XX WO200157270-A2.  
 XX 09-AUG-2001.  
 XX 29-JAN-2001; 2001WO-US00661.  
 XX 04-FEB-2000; 2000US-0180312.  
 XX 26-MAY-2000; 2000US-0207456.  
 XX 30-JUN-2000; 2000US-0608408.  
 XX 03-AUG-2000; 2000US-0632366.  
 XX 21-SEP-2000; 2000US-0234687.  
 XX 27-SEP-2000; 2000US-0236359.  
 XX 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-476286/51.  
 XX Novel single exon nucleic acid probe used to measuring gene expression  
 in a human breast  
 XX Claim 25; SEQ ID No 6049; 322pp; English.  
 XX The present invention relates to novel single exon nucleic acid probes.  
 XX The present sequence is one such probe. The probes are useful for  
 measuring human gene expression in a human breast sample, where the probe  
 hybridizes at high stringency to a nucleic acid expressed in the human  
 breast. The probes are useful for predicting, diagnosing, grading,  
 staging, monitoring and prognosing diseases of the human breast,  
 particularly those diseases with polygenic aetiology. The diseases  
 include: breast cancer, disorders of development, inflammatory diseases  
 of the breast, fibrocytic changes, proliferative breast disease and  
 non-carcinoma tumours.  
 XX Note: The sequence data for this patent did not form part of the printed  
 specification, but was obtained in electronic format directly from WIPO  
 at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 133 BP; 33 A; 44 C; 26 G; 30 T; 0 other;  
 XX  
 XX Query Match 53.9%; Score 19.4; DB 22; Length 133;  
 XX Best Local Similarity 79.3%; Pred. No. 63;  
 XX Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Qy 7 GCTACTGGTCTGCAAGGCAACTGACCTT 35  
 Db 46 gccactgcttctgcagtggactgcctt 74  
 RESULT 13  
 ID ABA17256 standard; DNA; 264 BP.  
 XX ABA17256;  
 XX AC  
 XX DT  
 XX 23-JAN-2002 (first entry)  
 XX Human nervous system related polynucleotide SEQ ID NO 9587.  
 XX Human; neutropic; neuroprotective; cytostatic; dermatological; virucide;  
 immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
 antiparkinsonian; antisickling; antianemic; antiarthritic; cancer;  
 antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
 antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;  
 antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
 neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
 XX Homo sapiens.  
 XX OS  
 XX WO200159063-A2.  
 XX PN

XX PN  
 XX WO200157270-A2.  
 XX 09-AUG-2001.  
 XX 29-JAN-2001; 2001WO-US00661.  
 XX 04-FEB-2000; 2000US-0180312.  
 XX 26-MAY-2000; 2000US-0207456.  
 XX 30-JUN-2000; 2000US-0608408.  
 XX 03-AUG-2000; 2000US-0632366.  
 XX 21-SEP-2000; 2000US-0234687.  
 XX 27-SEP-2000; 2000US-0236359.  
 XX 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488897/53.  
 XX Human genome-derived single exon nucleic acid probes useful for  
 analyzing gene expression in human placenta  
 XX Claim 25; SEQ ID No 14247; 654pp; English.  
 XX The present invention relates to single exon nucleic acid probes (SENP).  
 XX The present sequence is one such probe. The probes are useful for  
 producing a microarray for predicting, measuring and displaying gene  
 expression in samples derived from human placenta. The probes are useful  
 for antenatal diagnosis of human genetic disorders.  
 XX Sequence 133 BP; 33 A; 44 C; 26 G; 30 T; 0 other;  
 XX  
 XX Query Match 53.9%; Score 19.4; DB 22; Length 133;  
 XX Best Local Similarity 79.3%; Pred. No. 63;  
 XX Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Qy 7 GCTACTGGTCTGCAAGGCAACTGACCTT 35  
 Db 46 gccactgcttctgcagtggactgcctt 74  
 RESULT 12  
 ID AAI06058 standard; DNA; 133 BP.  
 XX AAI06058;  
 XX AC  
 XX DT  
 XX 09-OCT-2001 (first entry)  
 XX Probe #6049 used to measure gene expression in human breast sample.  
 XX Probe; human; breast disease; breast cancer; development disorder; ss;  
 inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
 XX Homo sapiens.  
 XX OS



Thu Jun 6 10:09:27 2002

us-08-881-509-7.rng

```

DR WPI; 2001-541565/60.
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases.
XX Disclosure: SEQ ID NO 9587; 1701pp + Sequence Listing; English.
XX The invention relates to novel genes (ABAI1004-ABAI2534) and proteins
XX (ABAI4678-ABAI8001) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemia;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 264 BP; 51 A; 65 C; 81 G; 67 T; 0 other;

Query Match 53.9%; Score 19.4; DB 22; Length 264;
Best Local Similarity 79.3%; Pred. No. 73;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 7 GCTACTGGTTCGACGCACTGACCTT 35
   || ||| ||||| | |||| |||
Db 54 GCCACTGCTTCGACGCACTGACCTT 26

RESULT 14
ABAI7254/c
ID ABAI7254 standard; DNA; 265 BP.
XX
XX ABAI7254;
XX
XX 23-JAN-2002 (first entry)
XX
XX Human nervous system related polynucleotide SEQ ID NO 9585.
XX
XX Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;
XX immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
XX antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
XX antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
XX antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
XX antiparasitic; cardiac; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX Homo sapiens.
XX
XX WO200159063-A2.
XX
XX 16-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01334.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
XX
XX 16-MAR-2000; 2000US-0189874.
XX
XX 17-MAR-2000; 2000US-0190076.
XX
XX 18-APR-2000; 2000US-0198123.
XX
XX 19-MAY-2000; 2000US-0205515.
XX
XX 07-JUN-2000; 2000US-0209457.
XX
XX 28-JUN-2000; 2000US-0214886.
XX
XX 30-JUN-2000; 2000US-0215135.
XX
XX 07-JUL-2000; 2000US-0216647.
XX
XX 11-JUL-2000; 2000US-0216880.
XX
XX 11-JUL-2000; 2000US-0217487.
XX
XX 14-JUL-2000; 2000US-0217496.
XX
XX 14-JUL-2000; 2000US-0218290.
XX
XX 26-JUL-2000; 2000US-0220963.
XX
XX 26-JUL-2000; 2000US-0220964.
XX
XX 14-AUG-2000; 2000US-0224518.
XX
XX 14-AUG-2000; 2000US-0225213.
XX
XX 14-AUG-2000; 2000US-0225214.
XX
XX 14-AUG-2000; 2000US-0225266.
XX
XX 14-AUG-2000; 2000US-0225267.
XX
XX 14-AUG-2000; 2000US-0225268.
XX
XX 14-AUG-2000; 2000US-0225270.
XX
XX 14-AUG-2000; 2000US-0225447.
XX
XX 14-AUG-2000; 2000US-0225757.
XX
XX 14-AUG-2000; 2000US-0225758.
XX
XX 14-AUG-2000; 2000US-0225759.
XX
XX 18-AUG-2000; 2000US-0226279.
XX
XX 22-AUG-2000; 2000US-0226681.
XX
XX 22-AUG-2000; 2000US-0226686.
XX
XX 22-AUG-2000; 2000US-0227182.
XX
XX 23-AUG-2000; 2000US-0227009.
XX
XX 30-AUG-2000; 2000US-0228924.
XX
XX 01-SEP-2000; 2000US-0229287.
XX
XX 01-SEP-2000; 2000US-0229343.
XX
XX 01-SEP-2000; 2000US-0229344.
XX
XX 01-SEP-2000; 2000US-0229345.
XX
XX 05-SEP-2000; 2000US-0229509.
XX
XX 05-SEP-2000; 2000US-0229513.
XX
XX 06-SEP-2000; 2000US-0230437.
XX
XX 06-SEP-2000; 2000US-0230438.
XX
XX 08-SEP-2000; 2000US-0231242.
XX
XX 08-SEP-2000; 2000US-0231243.
XX
XX 08-SEP-2000; 2000US-0231244.
XX
XX 08-SEP-2000; 2000US-0231413.
XX
XX 08-SEP-2000; 2000US-0231414.
XX
XX 08-SEP-2000; 2000US-0232080.
XX
XX 08-SEP-2000; 2000US-0232081.
XX
XX 12-SEP-2000; 2000US-0231968.
XX
XX 14-SEP-2000; 2000US-0232397.
XX
XX 14-SEP-2000; 2000US-0232398.
XX
XX 14-SEP-2000; 2000US-0232399.
XX
XX 14-SEP-2000; 2000US-0232400.
XX
XX 14-SEP-2000; 2000US-0232401.
XX
XX 14-SEP-2000; 2000US-0233063.
XX
XX 14-SEP-2000; 2000US-0233064.
XX
XX 21-SEP-2000; 2000US-0233065.
XX
XX 21-SEP-2000; 2000US-0234223.
XX
XX 21-SEP-2000; 2000US-0234274.
XX
XX 25-SEP-2000; 2000US-0234997.
XX
XX 25-SEP-2000; 2000US-0234998.
XX
XX 26-SEP-2000; 2000US-0235484.
XX
XX 27-SEP-2000; 2000US-0235834.
XX
XX 27-SEP-2000; 2000US-0235836.
XX
XX 29-SEP-2000; 2000US-0236327.
XX
XX 29-SEP-2000; 2000US-0236367.
XX
XX 29-SEP-2000; 2000US-0236368.
XX
XX 29-SEP-2000; 2000US-0236369.
XX
XX 29-SEP-2000; 2000US-0236370.
XX
XX 02-OCT-2000; 2000US-0236802.
XX
XX 02-OCT-2000; 2000US-0237037.
XX
XX 02-OCT-2000; 2000US-0237038.
XX
XX 02-OCT-2000; 2000US-0237039.
XX
XX 02-OCT-2000; 2000US-0237040.
XX
XX 13-OCT-2000; 2000US-0239935.
XX
XX 13-OCT-2000; 2000US-0239937.
XX
XX 20-OCT-2000; 2000US-0240960.
XX
XX 20-OCT-2000; 2000US-0241785.

```

PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 20-OCT-2000; 2000US-0242221.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251160.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 06-DEC-2000; 2000US-0256719.  
 PR 08-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 11-DEC-2000; 2000US-0251990.  
 PR 05-JAN-2001; 2001US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.

(HUMA\*) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-541565/60.

PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating nervous system  
 PT cancers and metastases -

XX Disclosure; SEQ ID NO 9585; 1701pp + Sequence Listing; English.

XX The invention relates to novel genes (AB11004-AB21534) and proteins

CC (AB114678-AB18001) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful

CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 265 BP; 51 A; 67 C; 81 G; 66 T; 0 other;

Query Match 53.9%; Score 19.4; DB 22; Length 265;  
 Best Local Similarity 79.3%; Pred. No. 73;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 7 GCTACTGGTTCTCGAAGCAACTGACCTT 35  
 || |||| |||| || |||| ||||  
 Db 54 GCCACTGCTTCTGCACTGGAACCTGCCCTT 26

RESULT 15

AB17255/c

ID AB17255 standard; DNA; 265 BP.

XX AC AB17255;

XX DT 23-JAN-2002 (first entry)

XX DE Human nervous system related polynucleotide SEQ ID NO 9586.

XX Human; nontropic; neuroprotective; cytostatic; dermatological; virucide;

KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;

KW antiparkinsonian; antisickling; antianemic; antiarthritic; cancer;

KW antirheumatic; hepatotropic; cerebroprotective; antinflammatory;

KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;

KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;

KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

XX Homo sapiens.

OS WO200159063-A2.

PN 16-AUG-2001.

PD 17-JAN-2001; 2001WO-US01334.

PF 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216847.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249246.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250391.
PR	01-DEC-2000;	2000US-0251160.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX	(HUMA-) HUMAN GENOME SCI INC.	
XX	Rosen CA, Barash SC, Ruben SM;	
XX	WPI; 2001-541565/60.	
XX	Nucleic acids encoding 3224 human nervous system antigen polypeptides,	
DR	useful for preventing, diagnosing and/or treating nervous system	
PT	cancers and metastases -	
PT	disclosure; SEQ ID NO 9586; 1701pp + Sequence Listing; English.	
XX	The invention relates to novel genes (ABAI1004-ABA21534) and proteins	
CC	(ABBI4678-ABBI8001) useful for preventing, treating or ameliorating	
CC	medical conditions e.g. by protein or gene therapy. The genes are	
CC	isolated from a range of human tissues disclosed in the specification.	
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful	
CC	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast	
CC	and ovarian cancer and other cancers of the adrenal gland, bone, bone	
CC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;	
CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune	
CC	haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's	
CC	disease, multiple sclerosis, rheumatoid arthritis and ulcerative	
CC	colitis; (c) cardiovascular disorders such as myocardial ischaemia;	
CC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and	
CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal	
CC	and parasitic infections.	
CC	Note: The sequence data for this patent did not form part of the	
CC	printed specification, but was obtained in electronic format directly	
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.	

Thu Jun '6 10:09:27 2002

us-08-881-509-7.rng

Page 11

XX  
SQ

Sequence 265 BP; 51 A; 67 C; 81 G; 66 T; 0 other;

Query Match

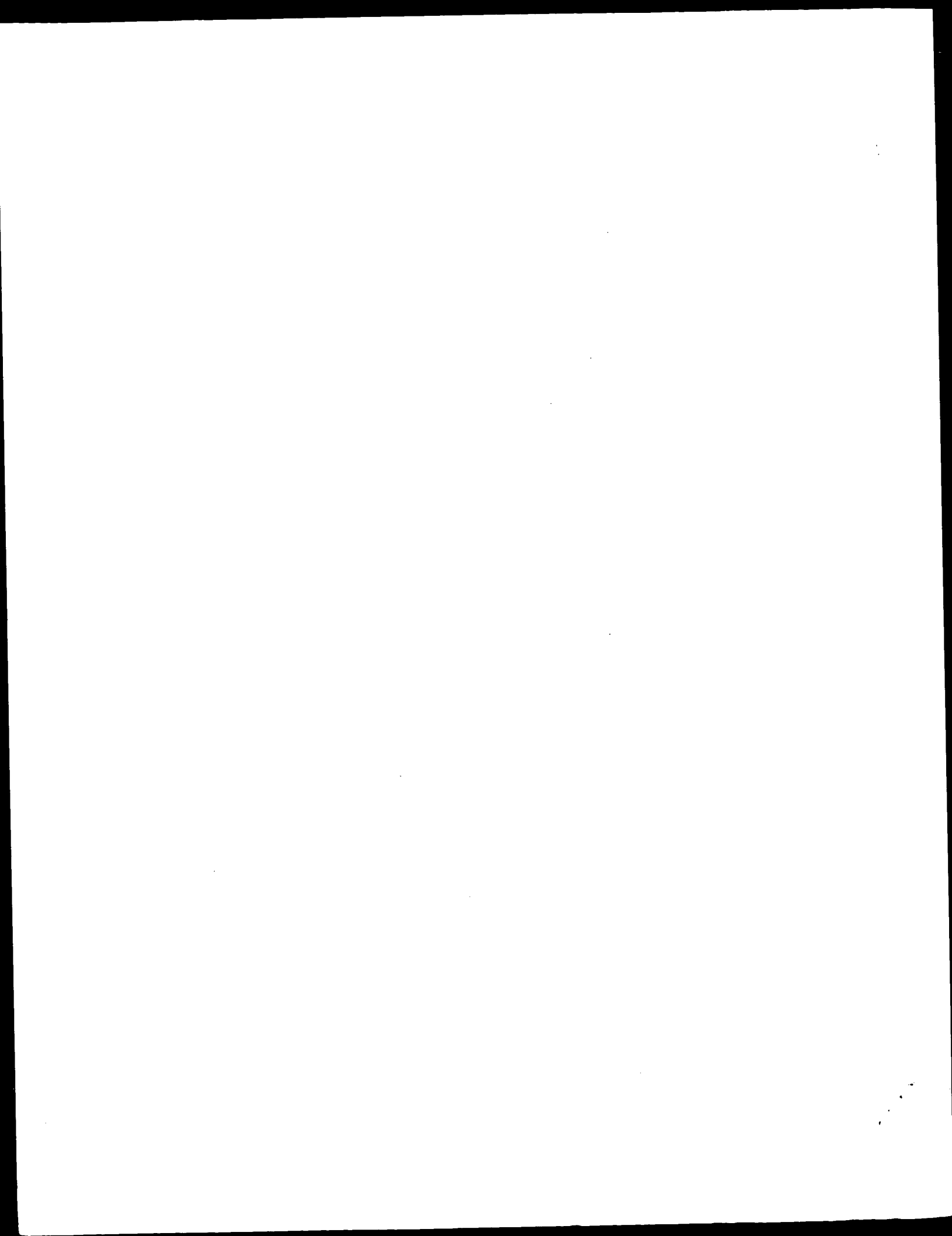
Best Local Similarity 53.9%; Score 19.4; DB 22; Length 265;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 7 GCTACTGTTCTGCAAGGCAACTGACCTT 35

Db 54 GCCACTGCTTCTGCAGTGAAGTGCCTT 26

Search completed: June 6, 2002, 07:00:24  
Job time: 3047 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: June 6, 2002, 07:38:15 ; Search time 78.65 Seconds  
(without alignments)  
112.432 Million cell updates/sec

Title: US-08-881-509-7  
Perfect score: 36  
Sequence: 1 TGCCTCGTACTGTTCTGCAAGCACTGACCTTT 36

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.4	53.9	15297	4	US-09-817-180-3
2	18.8	52.2	518	4	US-09-060-756-628
3	18.6	51.7	327	4	US-08-818-112-45
4	18.6	51.7	327	4	US-08-818-111-45
5	18.6	51.7	327	4	US-09-056-556-46
6	18.6	51.7	396	4	US-08-818-112-116
7	18.6	51.7	396	4	US-08-818-111-111
8	18.6	51.7	396	4	US-09-056-556-116
9	18.6	51.7	657	1	US-08-076-011-3
10	18.6	51.7	710	4	US-08-998-416-850
11	18.6	51.7	1616	4	US-08-818-112-112
12	18.6	51.7	1616	4	US-08-818-111-107
13	18.6	51.7	1616	4	US-09-056-556-112
14	18.6	51.7	7676	4	US-08-390-878-16
15	18.6	51.7	16885	1	US-08-076-011-1
16	18.6	51.7	17041	1	US-08-076-011-1
17	18.6	51.7	4403765	4	US-09-103-840A-2
18	18.6	51.7	4411529	4	US-09-103-840A-1
19	18.2	50.6	2220	6	5171850-1
20	18	50.0	2340	4	US-09-450-852-3
21	18	50.0	29604	3	US-08-781-891-207
22	17.8	49.4	1125	3	US-08-878-801-3
23	17.8	49.4	2388	4	US-09-276-531-89
24	17.6	48.9	213	4	US-08-905-223-46
25	17.6	48.9	1101	3	US-08-746-883-3
26	17.6	48.9	1447	4	US-09-180-437-3
27	17.6	48.9	2243	3	US-09-084-079-4

28	17.6	48.9	2404	4	US-09-180-437-4
29	17.4	48.3	6306	1	US-08-466-390-3
30	17.4	48.3	6306	1	US-08-470-950-3
31	17.4	48.3	6306	1	US-08-467-781-3
32	17.4	48.3	6306	1	US-08-195-487-3
33	17.4	48.3	6306	2	US-08-483-924-3
34	17.4	48.3	6306	5	PCT-US93-06160-3
35	17.4	48.3	13011	2	US-08-791-849A-14
36	17.4	48.3	49136	4	US-09-422-869-1
37	17.2	47.8	99	4	US-08-905-223-95
38	17.2	47.8	209	4	US-09-085-199B-31
39	17.2	47.8	1164	4	US-09-085-199B-1
40	17.2	47.8	1207	2	US-08-667-809B-1
41	17.2	47.8	1424	2	US-08-915-972A-21
42	17.2	47.8	1424	2	US-09-177-909-21
43	17.2	47.8	1424	2	US-09-056-105-21
44	17.2	47.8	3266	1	US-08-369-043-3
45	17.2	47.8	4406	1	US-08-369-043-1

## ALIGNMENTS

## RESULT 1

US-09-817-180-3/c  
; Sequence 3, Application US/09817180  
; Patent No. 6340584  
; GENERAL INFORMATION:  
; APPLICANT: GAN, Weiniu et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE REFERENCE: CL001183  
; CURRENT APPLICATION NUMBER: US/09/817,180  
; CURRENT FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 15297  
; TYPE: DNA  
; ORGANISM: Human  
US-09-817-180-3

Query Match 53.9%; Score 19.4; DB 4; Length 15297;  
Best Local Similarity 79.3%; Pred No. 22;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 5 TCGCTACTGTTCTGCAAGCACTGACC 33  
Db 9087 TCGTACAGGTTCTGCAGGGCAGGGACC 9059  
||||| ||||||| ||||| |||||

## RESULT 2

US-09-060-756-628  
; Sequence 628, Application US/09060756  
; Patent No. 6183957  
; GENERAL INFORMATION:  
; APPLICANT: Cole, Stewart  
; APPLICANT: Buchrieser-Brosch, Roland  
; APPLICANT: Gordon, Stephen  
; APPLICANT: Billault, Alain  
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM  
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA  
; FILE REFERENCE: 3495-0169  
; CURRENT APPLICATION NUMBER: US/09/060,756  
; CURRENT FILING DATE: 1998-04-16  
; NUMBER OF SEQ ID NOS: 743  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 628  
; LENGTH: 518  
; TYPE: DNA

us-08-881-509-7.rni

Thu Jun 6 10:09:27 2002

```

; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (various positions within the sequence)
; OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756-628

Query Match      52.2%; Score 18.8; DB 4; Length 518;
Best Local Similarity 74.2%; Pred. No. 18; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 8;

QY 2 GCCTCGCTACTGGTTCTGCAAGGCAACTGAC 32
    |||| |||| |||| |||| |||| |||| ||||
DB 467 gctggctacgggctcaacaaggcnaagac 497

RESULT 3
US-08-818-112-46/c
; Sequence 46, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-46

Query Match      51.7%; Score 18.6; DB 4; Length 327;
Best Local Similarity 72.7%; Pred. No. 19; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 9;

QY 2 GCCTCGCTACTGGTTCTGCAAGGCAACTGACCT 34
    |||| |||| |||| |||| |||| |||| ||||
DB 131 GCCGCGCACTGGCCCTGCAACGAACCTGCCGT 99

RESULT 5
US-09-056-556-46/c
; Sequence 46, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:

```



us-08-881-509-7.rni

Thu Jun 6 10:09:27 2002

```

; Sequence 116, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 116:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-116

; Query Match 51.7%; Score 18.6; DB 4; Length 396;
; Best Local Similarity 72.7%; Pred. No. 20;
; Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 GCCTCGCTACTGGTTCGCAAGGCAACTGACCT 34
Db 76 GCCGGCCACTGGCCCTGCAACGAACCTGCCGT 44

RESULT 9
US-08-076-011-3
; Sequence 3, Application US/08076011
; Patent No. 5521069
; GENERAL INFORMATION:
; APPLICANT: ONDA, Haruo
; APPLICANT: KIMURA, Chiharu
; APPLICANT: OHKUBO, Shoichi
; TITLE OF INVENTION: NOVEL DNA AND USE THEREOF
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN,
; ADDRESS: ROBERTS & CUSHMAN
; STREET: 130 Water Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:
; TELEFAX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 657 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-076-011-3

; Query Match 51.7%; Score 18.6; DB 1; Length 657;
; Best Local Similarity 72.7%; Pred. No. 23;
; Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 CTCGCTACTGGTTCGCAAGGCAACTGACCTTT 36
Db 485 CTCCTCTCTGGTTCGCGCTCTACAACTTTT 517

RESULT 10
US-08-998-416-850
; Sequence 850, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtie, Philipp
; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
; AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:
; TELEFAX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 657 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-076-011-3

```

; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 850:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 710 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: PAG1536UP  
; US-08-998-416-850

Query Match 51.7% Score 18.6; DB 4; Length 710;  
Best Local Similarity 72.7%; Pred. No. 23;  
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 CTCGCTACTGGTCTGCAAGGCAACTGACCTTT 36  
Db 7 CTCGGTACTGGAGAAACAAGGCTACTGTCCTT 39

RESULT 11  
US-08-818-112-112/c  
; Sequence 112, Application US/08818112  
; Patent No. 6290969  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; NUMBER OF SEQUENCES: 153  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/818,112  
; FILING DATE: 13-MAR-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C6  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 112:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1616 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-818-112-112

; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 850:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 710 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-998-416-850

Query Match 51.7% Score 18.6; DB 4; Length 1616;  
Best Local Similarity 72.7%; Pred. No. 29;  
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 GCCTCGCTACTGGTCTGCAAGGCAACTGACCT 34  
Db 1353 GCCGCGCACTGGCCCTGCAACGAACCTGCCGT 1321

RESULT 12  
US-08-818-111-107/c  
; Sequence 107, Application US/08818111  
; Patent No. 6338852  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; NUMBER OF SEQUENCES: 148  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/818,111  
; FILING DATE: 13-MAR-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C6  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 107:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1616 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-818-111-107

Query Match 51.7% Score 18.6; DB 4; Length 1616;  
Best Local Similarity 72.7%; Pred. No. 29;  
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 GCCTCGCTACTGGTCTGCAAGGCAACTGACCT 34  
Db 1353 GCCGCGCACTGGCCCTGCAACGAACCTGCCGT 1321

RESULT 13  
US-09-056-556-112/c  
; Sequence 112, Application US/09056556  
; Patent No. 6350456  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.

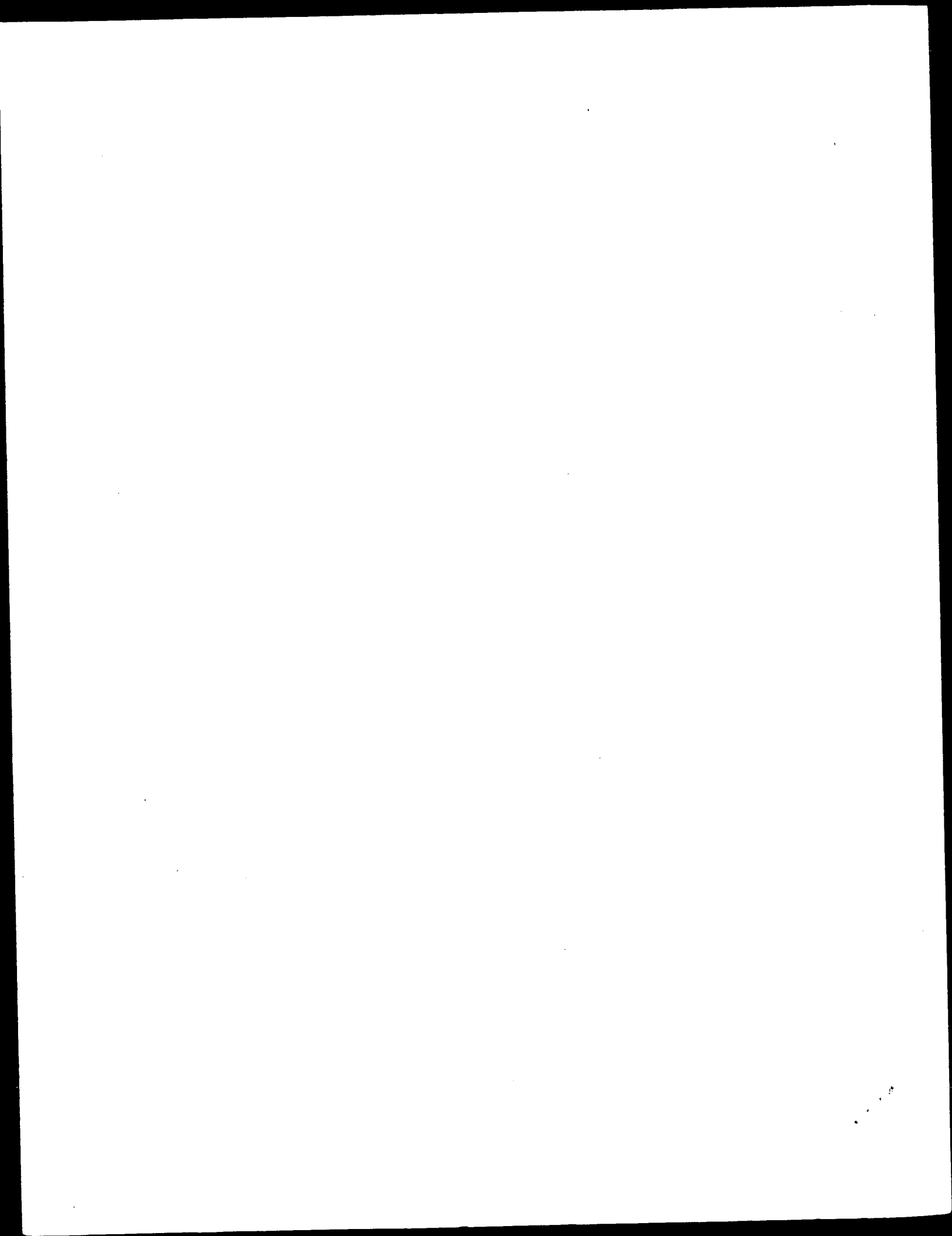
Db 4473 GCCGGCCACTGGCCCTGCAACGAACCTGCCGT 4441

Thu Jun 6 10:09:27 2002

us-08-881-509-7.rni

Page 7

Search completed: June 6, 2002, 07:38:20  
Job time: 5322 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:16:22 ; Search time 27.21 seconds  
(without alignments)  
42.377 Million cell updates/sec

Title: US-08-881-509-8  
Perfect score: 61  
Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR-71.\*

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	67.2	63	A1912	T-cell receptor al
2	40	65.6	384	A11843	hypothetical prote
3	39	63.9	82	S35770	T-cell receptor al
4	38	62.3	221	JC7307	endo-1,4-beta-xyla
5	38	62.3	394	G69230	endo-1,4-beta-gluc
6	37	60.7	71	E85976	probable galactosa
7	37	60.7	111	S57889	T cell receptor Er
8	37	60.7	245	A43181	short chain dehydr
9	37	60.7	251	A5104	probable galactosa
10	37	60.7	572	H85731	Rhs element associ
11	37	60.7	913	AH0989	hypothetical ABC t
12	37	60.7	1398	H90698	Rhs core protein
13	37	60.7	1398	B95549	hypothetical prote
14	37	60.7	1400	E90886	Rhs core protein
15	37	60.7	1426	H64780	rhd protein precu
16	36	59.0	72	AG2722	hypothetical prote
17	36	59.0	100	AG0447	probable phage-rel
18	36	59.0	357	D22337	UDP-N-acetylenolp
19	35	57.4	116	Q0EGR3	gene 44 protein -
20	35	57.4	363	WZB84	sporulation protei
21	35	57.4	423	B75503	myosin I beta, MWI
22	35	57.4	448	B45438	myosin-IC - mouse
23	35	57.4	807	D75634	myosin heavy chain
24	35	57.4	1028	S41749	myosin I heavy cha
25	35	57.4	1028	S37146	myosin I beta - hu
26	35	57.4	1028	A59253	probable histone d
27	35	57.4	1095	T13564	T-cell receptor (R
28	34	55.7	130	S00827	hypothetical prote
29	34	55.7	215	T00938	

30	34	55.7	300	2	S41171	transrepressor pro
31	34	55.7	335	2	D71275	hypothetical prote
32	34	55.7	354	2	E82850	fimbrial adhesin p
33	34	55.7	356	2	D82850	probable two-compo
34	34	55.7	422	2	E83083	hypothetical prote
35	34	55.7	445	2	T06051	orotate reductase
36	34	55.7	520	2	S72324	probable receptor -
37	34	55.7	646	2	B84852	nicotinic dehydroge
38	34	55.7	814	1	I39627	hypothetical prote
39	34	55.7	1145	1	B75625	probable PPE prote
40	33.5	54.9	350	2	H70929	T-cell receptor J-
41	33	54.1	223	2	S47194	probable transcript
42	33	54.1	261	2	A83196	hypothetical prote
43	33	54.1	270	2	G70812	hypothetical prote
44	33	54.1	271	2	B96773	site-specific DNA -
45	33	54.1	334	1	XYNHAL	

## ALIGNMENTS

RESULT 1  
A1912  
T-cell receptor alpha chain (D10) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 17-Mar-1999  
C:Accession: A41912  
R:Hong, S.C.; Chelouche, A.; Lin, R.; Shaywitz, D.; Braunstein, N.S.; Glimcher, L.; J  
Cell 69, 999-1009, 1992  
A>Title: An MHC interaction site maps to the amino-terminal half of the T cell recept  
A:Reference number: A41912; MUID:92298397  
A:Accession: A41912  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-63 <HON>  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: T-cell receptor

Query Match 67.2%; Score 41; DB 2; Length 63;  
Best Local Similarity 66.7%; Pred. No. 0.62;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLATGSARQLTF 12  
| | | | | : | | |  
Db 36 CAATGSFNKLT 47

RESULT 2  
A11843  
hypothetical protein alr0297 [imported] - Anabaena sp. (strain PCC 7120)  
C:Species: Anabaena sp.  
A>Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
C:Accession: A11843  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu  
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata  
DNA Res. 8, 203-213, 2001  
A>Title: Complete Genomic  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: A11843  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-384 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA077821.1; PID:g17135275; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr0297  
C:Superfamily: hypothetical protein MJ1598

Query Match 65.6%; Score 40; DB 2; Length 384;  
Best Local Similarity 63.6%; Pred. No. 5.4;

Thu Jun 6 10:09:40 2002

```

C:Accession: G69230
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwan,
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A:Reference number: A69000; MUID:98037514
A:Accession: G69230
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-394 <MTH>
A:Cross-references: GB:AE000871; GB:AE000666; NID:g2622069; PIDN:AAB85473.1; PID:g262
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH977
A:Start codon: TTG

Query Match 62.3%; Score 38; DB 2; Length 394;
Best Local Similarity 63.6%; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 LATGSARQLTF 12
Db 232 LATGETREVT 242

RESULT 6
E85976
probable galactosamine-6-phosphate isomerase agal_1 [Imported] - Escherichia coli (st
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85976
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimailanta, E.; Potamouisis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85976
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-71 <SPO>
A:Cross-references: GB:AE005174; NID:g12517742; PIDN:AAG58273.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: agal_1

Query Match 60.7%; Score 37; DB 2; Length 71;
Best Local Similarity 58.3%; Pred. No. 4.3;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLATGSARQLTF 12
Db 52 CLATGATPLLY 63

RESULT 7
S57889
T cell receptor Er3 alpha chain - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 23-Jul-1999
C:Accession: S57889
R:Giegerich, G.; Fette, M.; Meinel, E.; Epplen, J.T.; Wekerle, H.; Hinkkanen, A.
Eur. J. Immunol. 22, 753-758, 1992
A:Title: Diversity of T cell receptor alpha and beta chain genes expressed by human T
A:Reference number: S57889; MUID:92192091
A:Accession: S57889
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-111 <GIB>
A:Cross-references: EMBL:X57532; NID:g642982; PIDN:CAA40753.1; PID:g642983
C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Accession: G69230
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwan,
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A:Reference number: A69000; MUID:98037514
A:Accession: G69230
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-394 <MTH>
A:Cross-references: GB:AE000871; GB:AE000666; NID:g2622069; PIDN:AAB85473.1; PID:g262
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH977
A:Start codon: TTG

Query Match 63.9%; Score 39; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GSARQLTF 12
Db 60 GSARQLTF 67

RESULT 4
JC7307
endo-1,4-beta-xylanase (EC 3.2.1.8) - Penicillium sp.
N:Alternate names: endo-1,3-beta-xylanase; xylanase
C:Species: Penicillium sp.
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 02-Sep-2000
C:Accession: JC7307; PC7086
R:Kimura, T.; Ito, J.; Kawano, A.; Makino, T.; Kondo, H.; Karita, S.; Sakka, K.; Ohmiya,
BioSci. Biotechnol. Biochem. 64, 1230-1237, 2000
A:Title: Purification, characterization, and molecular cloning of acidophilic xylanase f
A:Reference number: JC7307
A:Accession: JC7307
A:Molecule type: DNA
A:Residues: 1-221 <KIM>
A:Cross-references: DDBJ:AB035540
A:Accession: PC7086
A:Molecule type: protein
A:Residues: 32-51 <KIT>
C:Genetics:
A:Gene: xynA
A:Introns: 89/2
C:Superfamily: endo-1,4-beta-xylanase; endo-1,4-beta-xylanase homology
C:Keywords: glycoprotein; glycosidase; hydrolase

Query Match 62.3%; Score 38; DB 2; Length 221;
Best Local Similarity 70.0%; Pred. No. 7.9;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATGSARQLTF 12
Db 84 STGSARDITF 93

RESULT 5
G69230,
endo-1,4-beta-glucanase related protein - Methanobacterium thermoautotrophicum (strain
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999

```

C:Keywords: T-cell receptor

Query Match 60.7%; Score 37; DB 2; Length 111;  
 Best Local Similarity 50.0%; Pred. No. 6.6;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 CLATGSARQLTF 12  
 ||||| |||  
 Db 88 CVCSTASKLTF 99

RESULT 8

short chain dehydrogenase Atu5181 [imported] - Agrobacterium tumefaciens (strain C58, D  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 01-Feb-2002  
 C:Accession: AH3181  
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, H  
 ; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 ster, E.W.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; PMID:11743193  
 A:Accession: AH3181  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-245 <KUR>  
 A:Cross-references: GB:AB008687; PIDN:AAL45870.1; PID:g17743613; GSPDB:GN00188  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: Atu5181  
 A:Genome: plasmid  
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match

Best Local Similarity 60.7%; Score 37; DB 2; Length 245;  
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 LATGSARQLT 11  
 ||||| |||  
 Db 224 LASGARQIT 233

RESULT 9

A65104  
 probable galactosamine-6-phosphate isomerase (EC 5.3.1.-) - Escherichia coli (strain K-1  
 C:Species: Escherichia coli  
 C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 16-Jul-1999  
 C:Accession: A65104  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
 .A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: A65104  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-251 <BLAT>  
 A:Cross-references: GB:AE000395; GB:U00096; NID:g1789524; PIDN:AAC76175.1; PID:g1789530;  
 A:Experimental source: strain K-12, substrain MGI655  
 C:Genetics:  
 A:Gene: agaI  
 C:Superfamily: glucosamine-6-phosphate isomerase  
 C:Keywords: intramolecular oxidoreductase; isomerase

Query Match

Best Local Similarity 60.7%; Score 37; DB 2; Length 251;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 CLATGSARQLTF 12  
 ||||| |||  
 Db 52 CLATGATPLITY 63

RESULT 10

H85731  
 Rhs element associated protein Z2259 [imported] - Escherichia coli (strain O157:H7, s  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: H85731  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
 iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: AB5480; MUID:21074935; PMID:11206551  
 A:Accession: H85731  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-572 <STO>  
 A:Cross-references: GB:AE005174; NID:g12515237; PIDN:AAG56316.1; GSPDB:GN00145; UWGP:  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: Z2259

Query Match

Best Local Similarity 60.7%; Score 37; DB 2; Length 572;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 ATGSARQLTF 12  
 ||||| |||  
 Db 389 ATGSTQMTW 398

RESULT 11

AH0989  
 hypothetical ABC transporter ATP-binding protein STY4224 [imported] - Salmonella ente  
 C:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A:Note: This species has also been called Salmonella typhi  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AH0989  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church  
 th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr  
 , S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AH0989  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-913 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD08044.1; PID:g16505025; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY4224  
 C:Superfamily: Escherichia coli probable ABC transporter yhiH; ATP-binding cassette h

Query Match

Best Local Similarity 60.7%; Score 37; DB 2; Length 913;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 LATGSARQL 10  
 ||||| |||  
 Db 227 LATGSAQQL 235

RESULT 12

H90698  
 Rhsd core protein with extension [imported] - Escherichia coli (strain O157:H7, subst  
 C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
 C:Accession: H90698  
 R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome  
 A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: H90698  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1398 <HAY>  
 A:CROSS-references: GB:BA000007; PIDN:BA033983.1; PID:gl3360018; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RMD 0509952  
 C:Genetics:  
 C:Superfamily: rhesF protein

Query Match 60.7%; Score 37; DB 2; Length 1400;  
 Best Local Similarity 70.0%; Pred. No. 71;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATGSARQLTF 12  
 |||||  
 Db 531 ATGSTROMTW 540

RESULT 15  
 H64780  
 rhesD protein precursor - Escherichia coli  
 C:Species: Escherichia coli  
 C>Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 24-Sep-1999  
 C:Accession: H64780; JS0625; B30092; I69401; S16026  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A>Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: H64780  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1426 <BLAT>  
 A:CROSS-references: GB:AE000156; GB:U00096; NID:gl786705; PIDN:AA073599.1; PID:gl7867  
 A:Experimental source: strain K-12, substrain MGL655  
 R:Sadosky, A.B.; Gray, J.A.; Hill, C.W.  
 Nucleic Acids Res. 19, 7177-7183, 1991  
 A>Title: The Rhes-E subfamily of Escherichia coli K-12.  
 A:Reference number: JS0625; MUID:92115567  
 A:Accession: JS0625  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-378, 'A', 380-1166, 'G', 1168-1426 <SAD>  
 A:CROSS-references: EMBL:X60999; NID:942732; PID:942733  
 A:Experimental source: strain K-12  
 R:Sadosky, A.B.; Davidson, A.; Lin, R.J.; Hill, C.W.  
 J. Bacteriol. 171, 636-642, 1989  
 A>Title: rhes gene family of Escherichia coli K-12.  
 A:Reference number: A91301; MUID:89123133  
 A:Accession: B30092  
 A:Molecule type: DNA  
 A:Residues: 1-100 <SA2>  
 A:CROSS-references: GB:M21764; GB:J04224; NID:gl47646; PIDN:AAA24542.1; PID:gl47649  
 R:Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehner, A.F.; Sadosky, A.B.; Vlazny, D.A.  
 J. Bacteriol. 172, 446-456, 1990  
 A>Title: Structure of the rhesA locus from Escherichia coli K-12 and comparison of rhes  
 A:Reference number: 154935; MUID:90094253  
 A:Accession: I69401  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1232-1426 <RES>  
 A:CROSS-references: GB:M29719; NID:gl47644; PIDN:AAA24541.1; PID:gl47645  
 C:Comment: the rhes core consist of two distinct parts: a large N-terminal core that i  
 C:Genetics:  
 C:Superfamily: rhesF protein  
 C:Keywords: transmembrane protein  
 F:1-26/Domain: signal sequence #status predicted <SIG>  
 F:27-1426/Product: rhesD protein #status predicted <MAT>  
 F:28-55/Domain: transmembrane #status predicted <TM>

C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
 C:Accession: H90698  
 R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome  
 A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: H90698  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1398 <HAY>  
 A:CROSS-references: GB:BA000007; PIDN:BA033983.1; PID:gl3360018; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RMD 0509952  
 C:Genetics:  
 C:Superfamily: rhesF protein

Query Match 60.7%; Score 37; DB 2; Length 1398;  
 Best Local Similarity 70.0%; Pred. No. 71;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATGSARQLTF 12  
 |||||  
 Db 531 ATGSTROMTW 540

RESULT 13  
 B85349  
 hypothetical protein Z0651 [imported] - Escherichia coli (strain O157:H7, substrain EDL9  
 C:Species: Escherichia coli  
 C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: B85349  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dialanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: B85349  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1398 <STO>  
 A:CROSS-references: GB:AE0051174; NID:gl2513390; PIDN:AA054854.1; GSPDB:GN00145; UWGP:Z06  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 C:Superfamily: rhesF protein

Query Match 60.7%; Score 37; DB 2; Length 1398;  
 Best Local Similarity 70.0%; Pred. No. 71;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATGSARQLTF 12  
 |||||  
 Db 531 ATGSTROMTW 540

RESULT 14  
 Rhes core protein with extension [imported] - Escherichia coli (strain O157:H7, substrain  
 C:Species: Escherichia coli  
 C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
 C:Accession: E90886  
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome  
 A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: E90886  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1400 <HAY>

JP

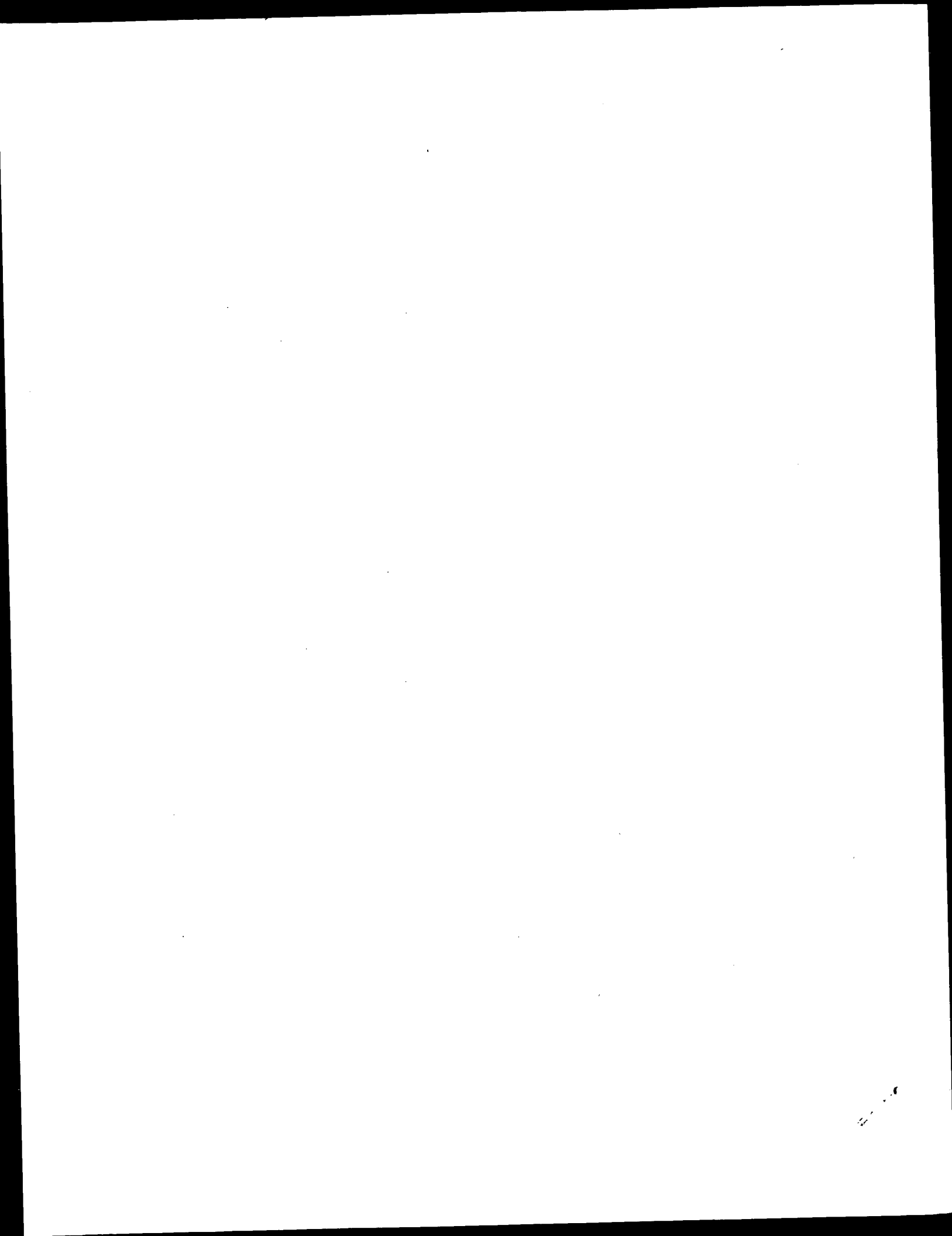
Thu Jun 6 10:09:40 2002

us-08-881-509-8.open.rpr

Page 5

Oy 3 ATGSARLTF 12  
|||||:  
Db 528 ATGSTROMTW 537

Search completed: June 6, 2002, 06:16:24  
Job time: 68 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:22:27 ; Search time 13.46 Seconds  
(without alignments)  
34.520 Million cell updates/sec

Title: US-08-881-509-8

Perfect score: 61

Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	62.3	189	1 MOBA_PASMU	P57971 pasteurella
2	37	60.7	251	1 AGAL_ECOLI	P42912 escherichia
3	37	60.7	1426	1 RHSD_ECOLI	P16919 escherichia
4	36	59.0	210	1 GL34_ARATH	Q9f1t3 arabidopsis
5	35	57.4	114	1 RT02_ASHGO	Q9hft3 ashbya goss
6	35	57.4	363	1 UL16_VZVD	P09293 varicella-z
7	35	57.4	1028	1 MY1C_HUMAN	O00159 homo sapien
8	35	57.4	1149	1 MY1C_MOUSE	Q9wt17 mus musculus
9	35	57.4	335	1 HD46_MOUSE	Q922v5 mus musculus
10	34	55.7	282	1 Y839_TREPA	O83811 treponema p
11	34	55.7	282	1 ZAN_RABIT	P57999 oryctolagus
12	33	54.1	315	1 SX22_HUMAN	O15370 homo sapien
13	33	54.1	334	1 MTN3_NEILA	P24582 neisseria l
14	33	54.1	336	1 G3P_CORGL	Q01651 corynebacte
15	33	54.1	357	1 PMFE_PROMI	P35222 proteus mir
16	33	54.1	363	1 PGL1_COLLN	Q00446 collettotric
17	33	54.1	380	1 PGL2_PENOL	Q9y833 penicillium
18	33	54.1	458	1 RXRB_RAT	P49743 rattus norv
19	33	54.1	496	1 ERG1_YEAST	P32476 saccharomyc
20	33	54.1	520	1 RXRB_MOUSE	P28704 mus musculus
21	33	54.1	533	1 RXRB_HUMAN	P28702 homo sapien
22	33	54.1	673	1 FLID_TREMA	Q9kww7 treponema m
23	33	54.1	894	1 YH1H_ECOLI	P37624 escherichia
24	33	54.1	1505	1 PK3G_RAT	O70173 rattus norv
25	33	54.1	2813	1 VNF_CANFA	Q28295 canis fami
26	32	52.5	107	1 YB11_BACHD	O9kdv0 bacillus ha
27	32	52.5	216	1 UPP_FORGI	P27710 cryptococcu
28	32	52.5	312	1 TRPF_CRYNE	P77980 salmonella
29	32	52.5	362	1 AROB_SALTY	P41749 aspergillus
30	32	52.5	363	1 PGLA_ASPFL	P35335 aspergillus
31	32	52.5	363	1 PGLR_ASPOR	P49575 aspergillus
32	32	52.5	363	1 PGLR_ASPFA	P18503 ephydatia m
33	32	52.5	366	1 CAS4_EPHMU	

34 32 52.5 496 1 E1BL\_ADE05 P03243 human adeno  
35 32 52.5 534 1 CN9A\_MOUSE O70628 mus musculu  
36 32 52.5 556 1 RORA\_HUMAN P35398 homo sapien  
37 32 52.5 564 1 BMRP\_CANAL P28873 candida alb  
38 32 52.5 610 1 YQ5C\_CAEEL O76083 homo sapien  
39 32 52.5 705 1 STT3\_HUMAN O09466 caenorhabdi  
40 32 52.5 811 1 Y104\_SYNV3 P49777 homo sapien  
41 32 52.5 1156 1 GLH4\_CAEEL P54371 synechocyst  
42 32 52.5 1873 1 CCAS\_HUMAN O76743 caenorhabdi  
43 32 52.5 2482 1 VWF\_PIG Q13698 homo sapien  
44 32 52.5 2700 1 ZAN\_HUMAN Q28833 sus scrofa  
45 32 52.5 2700 1 ZAN\_HUMAN Q9y493 homo sapien

## ALIGNMENTS

RESULT 1  
MOBA\_PASMU  
ID MOBA\_PASMU STANDARD; PRT; 189 AA.  
AC P57971;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE Probable molybdopterin-guanine dinucleotide biosynthesis protein A.  
GN MOBA OR PM1799.  
OS Pasteurella multocida.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Pasteurella.  
OX NCBI\_TaxID=747;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=PM70;  
RC MEDLINE=21145866; PubMed=11248100;  
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
RT "Complete genomic sequence of Pasteurella multocida Pm70.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
CC -!- FUNCTION: LINKS A GUANOSINE 5'-PHOSPHATE TO MOLYBDOTERIN (MPT)  
FORMING MOLYBDOTERIN GUANINE DINUCLEOTIDE (MGD) (BY SIMILARITY).  
CC -!- PATHWAY: MOLYBDENUM COFACTOR BIOSYNTHESIS.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE MOBA FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; AF006217; AAK03883.1; -  
DR HSSP; P32173; 1E5K.  
KW Molybdenum cofactor biosynthesis; GTP-binding; Complete proteome.  
SQ SEQUENCE 189 AA; 20922 MW; A467B28E1F42C2C7 CRC64;

Query Match 62.3%; Score 38; DB 1; Length 189;  
Best Local Similarity 72.7%; Pred. No. 2.3;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 LATGSARQLTF 12  
|||||  
Db 150 LATGERRMLTF 160

RESULT 2  
AGAL\_ECOLI  
ID AGAL\_ECOLI STANDARD; PRT; 251 AA.  
AC P42912;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)





```

DR EMBL; AE000156; AAC73599.1;
DR EMBL; U82664; AAB40251.1;
DR EMBL; AF044500; AAC32467.1;
DR EMBL; X60999; CAA43314.1;
DR EMBL; M21764; AAA24542.1;
DR PIR; B30092; B30092;
DR PIR; JS0625; JS0625;
DR PIR; S16026; S16026;
DR EcoGene; EG10849; rhdS;
DR InterPro; IPR001826; Rhs;
DR PRINTS; PR00394; RhsPROTEIN;
KW Signal; Multigene family; Transmembrane; Repeat; Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 1426
FT TRANSMEM 28 55
FT CONFLICT 379 379
FT CONFLICT 1167 1167
FT SEQUENCE 1426 AA; 159724 MW; 122D697AA449BC2F CRC64;

Query Match 60.7%; Score 37; DB 1; Length 1426;
Best Local Similarity 70.0%; Pred. No. 27;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATGSGARQLTF 12
    ||||| ||:|
DB 528 ATGSTRQMTW 537

RESULT 4
GL34_ARATH
ID GL34_ARATH STANDARD; PRT; 210 AA.
AC Q9FLT3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Putative germin-like protein subfamily 3 member 4 precursor.
GN AT5G61750 OR MAC9.6 OR MAC9_50.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016721; PubMed=11130714;
RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,
RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,
RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA Kirchhoff K., Toth K., King L., Bahret A., Miller B., Barra M.,
RA Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
RA Entian K.-D., Terry N., Hartley N., Bent E., Johnson S.,
RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
RA Ransperger U., Wedler H., Balke K., Wedler E., Peters S.,
RA Van Staveren M., Dirke W., Mooijman P., Klein Lankhorst R.,
RA Weltzienegger T., Bothe G., Rose M., Hauf J., Berner S., Hempel S.,
RA Feldpausch M., Lamberth S., Villarroel R., Giesen J., Ardies W.,
RA Bents O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schoof H.,
RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Fransch P.F.;
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
thaliana.";
```

```

RL Nature 408:823-826(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=98290546; PubMed=9628582;
RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
Sequence features of the regions of 1,456,315 bp covered by nineteen
physically assigned pl and TAC clones.";
RL DNA Res. 5:41-54(1998).
CC -!- FUNCTION: MAY PLAY A ROLE IN PLANT DEFENSE. HAS PROBABLY NO
CC OXALATE OXIDASE ACTIVITY EVEN IF THE ACTIVE SITE IS CONSERVED.
CC -!- SUBUNIT: OLIGOMER (BELIEVED TO BE A PENTAMER BUT PROBABLY
CC HEXAMER) (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Apoplast (By similarity).
CC -!- SIMILARITY: BELONGS TO THE GERMIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB010069; BAB10075.1;
DR InterPro; IPR001929; Germin.
DR Pfam; PF01072; Germin; 1.
DR PRINTS; PR00325; GERMIN.
DR PROSITE; PS00725; GERMIN; 1.
KW Apoplast; Cell wall; Signal; Glycoprotein; Manganese; Metal-binding;
KW Multigene family; Hypothetical protein.
FT SIGNAL 1 18
FT CHAIN 19 210
FT METAL 106 106
FT METAL 108 108
FT METAL 113 113
FT METAL 152 152
FT DISULFID 27 44
FT CARBOHYD 73 73
FT SEQUENCE 210 AA; 22890 MW; F80695370A6F5284 CRC64;

Query Match 59.0%; Score 36; DB 1; Length 210;
Best Local Similarity 63.6%; Pred. No. 6.3;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CLATGSARQLT 11
    ||||| :|
DB 195 CLATGSGCRVT 205

RESULT 5
RT02_ASHGO
ID RT02_ASHGO STANDARD; PRT; 114 AA.
AC Q9HF53;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mitochondrial 40S ribosomal protein MRP2.
GN MRP2.
OS Ashbya gossypii (Eremothecium gossypii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Eremotheciales; Eremothecium.
OX NCBI_TaxID=33169;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016721; PubMed=11130714;
RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,
RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,
RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA Kirchhoff K., Toth K., King L., Bahret A., Miller B., Barra M.,
RA Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
RA Entian K.-D., Terry N., Hartley N., Bent E., Johnson S.,
RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
RA Ransperger U., Wedler H., Balke K., Wedler E., Peters S.,
RA Van Staveren M., Dirke W., Mooijman P., Klein Lankhorst R.,
RA Weltzienegger T., Bothe G., Rose M., Hauf J., Berner S., Hempel S.,
RA Feldpausch M., Lamberth S., Villarroel R., Giesen J., Ardies W.,
RA Bents O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schoof H.,
RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Fransch P.F.;
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
thaliana.";
```

Thu Jun 6 10:09:41 2002

CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE S14P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF210628; AAC41250.1; -  
 CC InterPro: IPR001209; Ribosomal\_S14.  
 CC Pfam: PF00253; Ribosomal\_S14; 1.  
 CC PROSITE: PS00527; RIBOSOMAL\_S14; 1.  
 CC Ribosomal protein; Mitochondrion.  
 CC KW RIBOSOMAL protein; Mitochondrion.  
 CC SEQUENCE 114 AA; 13017 MW; BD15BFS5A2107F02F CRC64;  
 CC SQ SEQUENCE 114 AA; 13017 MW; BD15BFS5A2107F02F CRC64;  
 CC -----  
 CC Query Match 57.4%; Score 35; DB 1; Length 114;  
 CC Best Local Similarity 60.08; Pred. No. 5.4;  
 CC Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 CC -----  
 CC QY 1 CLATGSARQLT 10  
 CC :|||||:  
 CC Db 77 CIATGHARSV 86  
 CC -----  
 CC RESULT 6  
 CC UL16\_VZVD STANDARD; PRT; 363 AA.  
 CC ID UL16\_VZVD  
 CC AC P09293;  
 CC DT 01-MAR-1989 (Rel. 10, Created)  
 CC DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 CC DT 01-APR-1993 (Rel. 25, Last annotation update)  
 CC DE Gene 44 protein.  
 CC DE 44.  
 CC OS Varicella-zoster virus (strain Dumas) (VZV).  
 CC GN Varicella-zoster virus; dsDNA viruses, no RNA stage; Herpesviridae;  
 CC OC Alphaherpesvirinae; Varicelloviruses.  
 CC OX NCBI\_TaxID=10338;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RA Davison A.J., Scott J.E.;  
 CC RT "The complete DNA sequence of varicella-zoster virus."  
 CC RL J. Gen. Virol. 67:1759-1816(1986).  
 CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL16,  
 CC HSV-6 ORF11R, EBV-1 46, HCMV UL94, EBV BGLP2, HSV 33, AND VZV 44.  
 CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X04370; CAA27927.1; -  
 CC PIR: I27341; WZBEA4  
 CC InterPro: IPR004286; UL16\_UL94.  
 CC Pfam: PF03044; UL16\_UL94; 1.  
 CC SEQUENCE 363 AA; 40244 MW; 5F78F627E6261B87 CRC64;  
 CC SQ SEQUENCE 363 AA; 40244 MW; 5F78F627E6261B87 CRC64;  
 CC -----  
 CC Query Match 57.4%; Score 35; DB 1; Length 363;  
 CC Best Local Similarity 77.8%; Pred. No. 17;  
 CC Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 CC -----  
 CC QY 3 ATGSARQLT 11  
 CC :|||||:  
 CC Db 12 ATGAARKLT 20  
 CC -----

RESULT 7  
 MYIC\_HUMAN STANDARD; PRT; 1028 AA.  
 ID MYIC\_HUMAN  
 AC O00159;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myosin Ic (Myosin I beta) (MMI-beta) (MMIb).  
 DE MYO1C.  
 GN MYO1C.  
 OS Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=97237053; PubMed=9119401;  
 RA Crozet F., Fizes C., Levi-Acobas F., Depetris D., Mattei M.-G.,  
 RA Hamel D., Pujol R., Petit C.;  
 RA "Cloning of the genes encoding two murine and human cochlear  
 RT unconventional type I myosins."  
 RL Genomics 40:332-341(1997).  
 CC -!- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE  
 CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.  
 CC THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS  
 CC COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY  
 CC SIMILARITY).  
 CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 2 IQ DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X98507; CAA67131.1; -  
 CC DR HSSP; P08799; LLVK.  
 CC DR MIM; 606538;  
 CC DR InterPro: IPR000048; IQ.  
 CC DR InterPro: IPR001609; myosin\_head.  
 CC DR Pfam: PF00612; IQ; 3.  
 CC DR Pfam: PF00063; myosin\_head; 1.  
 CC DR PRINTS; PR00193; MYOSINHEAVY.  
 CC DR PRODOM; PD000355; myosin\_head; 1.  
 CC DR SMART; SM00015; IQ; 2.  
 CC DR SMART; SM00242; MYSC; 1.  
 CC DR PROSITE; PS00096; IQ; 2.  
 CC KW Myosin; ATP-binding; Actin-binding; Calmodulin-binding; Repeat;  
 CC Multigene family.  
 CC FT DOMAIN 1 683 MYOSIN HEAD-LIKE.  
 CC FT DOMAIN 699 722 IQ 1.  
 CC FT DOMAIN 723 751 IQ 2.  
 CC FT NP\_BIND 105 112 ATP (POTENTIAL).  
 CC SEQUENCE 1028 AA; 118037 MW; 0E9C3680527F85C6 CRC64;  
 CC SQ SEQUENCE 1028 AA; 118037 MW; 0E9C3680527F85C6 CRC64;  
 CC -----  
 CC Query Match 57.4%; Score 35; DB 1; Length 1028;  
 CC Best Local Similarity 58.3%; Pred. No. 49;  
 CC Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 CC -----  
 CC QY 1 CLATGSARQLT 12  
 CC :|||||:  
 CC Db 463 CLRPGCATDLAT 474  
 CC -----  
 CC RESULT 8  
 CC MYIC\_MOUSE STANDARD; PRT; 1028 AA.  
 CC ID MYIC\_MOUSE

AC Q9W17: O08834; O08571; O9QW54;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myosin Ic (Myosin I beta) (MMIB).  
 GN MYOIC.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS A AND B).  
 RC STRAIN=OBA/2J; TISSUE=Brain;  
 RX MEDLINE=97325785; PubMed=9182797;  
 RA Hamilton B.A., Smith D.J., Mueller K.L., Kerrebrock A.W.,  
 RA Bronson R.T., van Berkel V., Daly M.J., Kruglyak L., Reeve M.P.,  
 RA Nemausier J.L., Hawkins T.L., Rubin E.M., Lander E.S.;  
 RT "The vibrator mutation causes neurodegeneration via reduced expression  
 RT of PTP alpha: positional complementation cloning and extragenic  
 RL suppression.";  
 RL Neuron 18:711-722(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM A).  
 RC STRAIN=C57BL/6J; TISSUE=Lung;  
 RX MEDLINE=21083660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,  
 RA Saito T., Okazaki Y., Gojbori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staebli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE OF 1-807 FROM N.A. (ISOFORM B).  
 RC STRAIN=BALB/C; TISSUE=Cochlea;  
 RX MEDLINE=97237053; PubMed=9119401;  
 RA Crozet F., Anraoui A.E., Blanchard S., Lenoir M., Ripoll C., Vago P.,  
 RA Hamel C., Fzames C., Levi-Acobas F., Depetris D., Mattei M.-G.,  
 RA Weil D., Pujol R., Petit C.;  
 RT "Cloning of the genes encoding two murine and human cochlear  
 RT unconventional type I myosins.";  
 RL Genomics 40:332-341(1997).  
 RN [4]  
 RP SEQUENCE OF 146-593 FROM N.A.  
 RX MEDLINE=93194946; PubMed=8449986;  
 RA Sherr E.H., Joyce M.P., Greene L.A.;  
 RA "Mammalian myosin I alpha, I beta, and I gamma: new widely expressed  
 RL genes of the myosin I family.";  
 RL J. Cell Biol. 120:1405-1416(1993).  
 CC -I- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE  
 CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.  
 CC THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS  
 CC COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY  
 CC SIMILARITY).  
 CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -I- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC -I- SIMILARITY: CONTAINS 2 IQ DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL: U96723; AAC53264.1; -;  
 DR EMBL: AK004743; BAB23524.1; -;  
 DR EMBL: U96726; AAC60758.1; -;  
 DR EMBL: X96638; CAA67856.1; -;  
 DR HSSP: P08799; ILVK.  
 DR MGI: 106612; Myo1c.  
 DR InterPro: IPR000048; IQ.  
 DR InterPro: IPR001609; myosin\_head.  
 DR Pfam: PF00612; IQ; 3.  
 DR Pfam: PF00063; myosin\_head; 1.  
 DR PRINTS: PR00193; MYOSINHEAVY.  
 DR ProDom: PD000355; myosin\_head; 1.  
 DR SMART: SM00015; IQ; 2.  
 DR SMART: SM00242; MYSC; 1.  
 DR PROSITE: PS50096; IQ; 2.  
 DR Myosin; ATP-binding; Actin-binding; Calmodulin-binding; Repeat;  
 KW Multigene family; Alternative splicing.  
 FT DOMAIN 1 683 MYOSIN HEAD-LIKE.  
 FT DOMAIN 699 722 IQ 1.  
 FT DOMAIN 723 751 IQ 2.  
 FT NP\_BIND 105 112  
 FT VARSPLIC 897 1028  
 FT  
 FT CONFLICT 34 35  
 FT CONFLICT 62 64  
 FT CONFLICT 216 216  
 FT CONFLICT 237 237  
 FT CONFLICT 251 252  
 FT CONFLICT 353 353  
 FT CONFLICT 366 368  
 FT CONFLICT 411 411  
 FT CONFLICT 482 484  
 FT CONFLICT 543 543  
 FT CONFLICT 572 572  
 FT CONFLICT 700 700  
 FT CONFLICT 786 786  
 FT CONFLICT 807 807  
 SQ SEQUENCE 1028 AA; 118155 MW; 2CF5535D816884DF CRC64;  
 RR -> GG (IN REF. 3).  
 SRQ -> RRK (IN REF. 3).  
 T -> A (IN REF. 4).  
 C -> F (IN REF. 4).  
 VM -> LL (IN REF. 4).  
 R -> A (IN REF. 4).  
 LAS -> VPA (IN REF. 4).  
 Q -> R (IN REF. 4).  
 VKP -> IKH (IN REF. 4).  
 M -> T (IN REF. 4).  
 S -> G (IN REF. 4).  
 Q -> R (IN REF. 3).  
 R -> G (IN REF. 3).  
 E -> D (IN REF. 3).  
 Query Match 57.4%; Score 35; DB 1; Length 1028;  
 Best Local Similarity 58.3%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 CLATGSAROLTF 12  
 Db 463 CLRPGATDLTF 474  
 RESULT 9  
 ID HDA6\_MOUSE STANDARD; PRT; 1149 AA.  
 AC Q922V5;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Histone deacetylase 6 (HD6) (Histone deacetylase mHDA2).  
 GN HDAC6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]

Thu Jun 6 10:09:41 2002

```

SEQUENCE FROM N.A.
STRAIN=C57BL/6J; TISSUE=Petal;
MEDLINE=99107904; PubMed=9891014;
Verdel A., Khochbin S.;
"Identification of a new family of higher eukaryotic histone
deacetylases. Coordinate expression of differentiation-dependent
chromatin modifiers."
J. Biol. Chem. 274:2440-2445(1999).
-!- FUNCTION: RESPONSIBLE FOR THE DEACETYLATION OF LYSINE RESIDUES ON
THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4).
HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL
REGULATION, CELL CYCLE PROGRESSION AND DEVELOPMENTAL EVENTS (BY
SIMILARITY).
-!- SUBCELLULAR LOCATION: Nuclear (By similarity).
-!- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
FAMILY. HD SUBFAMILY 2.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL; AF006603; AAD09835.2; -
MGD; MGI:1333752; Hdac6.
InterPro: IPR00286; Hist_deacetylase.
InterPro: IPR001607; zf-UBP.
DR Pfam; PF00850; Hist_deacetyl; 2.
DR Pfam; PF02148; zf-UBP; 1.
DR PRINTS; PR01270; HDASUPER.
DR SMART; SM00290; ZNF_UBP; 1.
DR Hydrolase; Nuclear protein; Repeat.
KW DOMAIN 87 403 HISTONE DEACETYLASE 1.
FT DOMAIN 481 799 HISTONE DEACETYLASE 2.
FT DOMAIN 455 460 POLY-GLU.
SQ SEQUENCE 1149 AA; 125703 MW; 2B98CDB282CE0D1D CRC64;

Query Match 57.4%; Score 35; DB 1; Length 1149;
Best Local Similarity 70.0%; Pred. No. 55;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CLATGSARQL 10
| | | | | | | |
DB 184 CLATGSVLRL 193

RESULT 10
Y839_TREPA STANDARD; PRT; 335 AA.
AC O83811;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical lipoprotein TP0839 precursor.
GN TP0839.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=96659876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artlach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RA "Complete genome sequence of Treponema pallidum, the syphilis

```

```

spirochete."
Science 281:375-388(1998).
-!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
(Potential).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL; AE001254; AAC65809.1; -
TIGR; TP0839; -
PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
KW Hypothetical protein; Membrane; Lipoprotein; Signal;
KW Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 335 HYPOTHETICAL LIPOPROTEIN TP0839.
FT LIPID 22 22 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 335 AA; 37364 MW; 2F4C5FE154DAB869 CRC64;

Query Match 55.7%; Score 34; DB 1; Length 335;
Best Local Similarity 60.0%; Pred. No. 25;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CLATGSARQL 10
| | | | | | | |
DB 311 CLSTGGSARQL 320

RESULT 11
ZAN_RABIT STANDARD; PRT; 2282 AA.
ID ZAN_RABIT
AC P57999;
DT 16-OCT-2001 (Rel. 40, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Zonadhesin (Fragment).
ZAN.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Lea I.A., Sivashanmugam P., Richardson R.T., O'Rand M.G.;
RT "Sequence of rabbit zonadhesin."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
SIGNALING (BY SIMILARITY).
CC -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE 1 MEMBRANE PROTEIN, EXCLUSIVELY ON THE
APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
CC -!- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
ZONA PELLUCIDA.
CC -!- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC -!- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC -!- SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

```



Db 100 CLGSGSERQ 108

RESULT 14

G3P\_CORGL STANDARD; PRT; 336 AA.

ID G3P\_CORGL

AC 001651:

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH).

GN GAP.

OS Corynebacterium glutamicum (Brevibacterium flavum).

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriaceae;

OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;

OC Corynebacterium.

OX NCBI\_TaxID=1718;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 13059 / AS019;

RX MEDLINE=93015645; PubMed=1400158;

RA Rikmanns B.J.:

RT "Identification, sequence analysis, and expression of a

RT Corynebacterium glutamicum gene cluster encoding the three glycolytic

RT enzymes glyceraldehyde-3-phosphate dehydrogenase, 3-phosphoglycerate

RT kinase, and triosephosphate isomerase."

RL J. Bacteriol. 174:6076-6086(1992).

CC -|- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate +

CC NAD(+) = 3-phospho-D-glyceroyl phosphate + NADH.

CC -|- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.

CC -|- SUBUNIT: HOMOTETRAMER.

CC -|- SUBCELLULAR LOCATION: Cytoplasmic.

CC -|- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE

CC -|- DEHYDROGENASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; X59403; CAA42045.1; -

CC PIR; S23910; S23910.

CC PIR; A43260; A43260.

CC HSP; P00362; ICD1.

CC InterPro; IPR000173; GAP\_DH.

CC Pfam; PF00044; gpdh\_1.

CC Pfam; PF02800; gpdh\_C; 1.

CC PRINTS; PR00078; G3PDHGRNASE.

CC PROSITE; PS00071; GAPDH; 1.

CC Glycolysis; 153 153

CC BINDING 180 180

CC ACT\_SITE 180 180

CC SEQUENCE 336 AA; 36199 MW; 5514A0A0CF078219 CRC64;

DR EMBL; X59403; CAA42045.1; -

DR PIR; S23910; S23910.

DR PIR; A43260; A43260.

DR HSP; P00362; ICD1.

DR InterPro; IPR000173; GAP\_DH.

DR Pfam; PF00044; gpdh\_1.

DR Pfam; PF02800; gpdh\_C; 1.

DR PRINTS; PR00078; G3PDHGRNASE.

DR PROSITE; PS00071; GAPDH; 1.

KW Glycolysis; 153 153

FT BINDING 180 180

FT ACT\_SITE 180 180

FT SEQUENCE 336 AA; 36199 MW; 5514A0A0CF078219 CRC64;

SQ

Query Match 54.1%; Score 33; DB 1; Length 336;

Best Local Similarity 77.8%; Pred. No. 40;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 TGSARQLTF 12

Db 239 TGSATDLTF 247

RESULT 15

PMFE\_PROMI STANDARD; PRT; 357 AA.

ID PMFE\_PROMI

AC P53522;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

QY 1 CLATGSARQ 9

Db 142 CLAAGAAEQ 150

Query Match 54.1%; Score 33; DB 1; Length 334;

Best Local Similarity 66.7%; Pred. No. 40;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CLATGSARQ 9

Db 142 CLAAGAAEQ 150

RESULT 13

MTN3\_NEILA STANDARD; PRT; 334 AA.

ID MTN3\_NEILA

AC P24582;

DT 01-MAR-1992 (Rel. 21, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Modification methylase NlaIII (EC 2.1.1.72) (Adenine-specific

DE methyltransferase NlaIII) (M.NlaIII).

GN NLAIII.

OS Neisseria lactamica.

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

OX NCBI\_TaxID=486;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 23970 / NRCC 2118;

RX MEDLINE=91117164; PubMed=2277628;

RA Labbe D., Hoelke H.J., Lau P.C.K.;

RT "Cloning and characterization of two tandemly arranged DNA

RT methyltransferase genes of Neisseria lactamica: an adenine-specific

RT M.NlaIII and a cytosine-type methylase."

RL Mol. Gen. Genet. 224:101-110(1990).

CC -|- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE

CC CATG, CAUSES SPECIFIC METHYLATION ON A-2 ON BOTH STRANDS AND

CC PROTECTS THE DNA FROM CLEAVAGE BY THE NLAIII ENDONUCLEASE.

CC -|- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine - S-

CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.

CC -|- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-57 IS THE INITIATOR.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; X54485; CAA38356.1; -

CC PIR; S12036; XYNHAL.

CC REBASE; 3468; M.NlaIII.

CC InterPro; IPR002294; D12N6\_mtfase.

CC InterPro; IPR002052; N6\_Mtase.

CC Pfam; PF02086; MethyltransferaseD12; 1.

CC PRINTS; PR00505; D12N6MTFRASE.

CC PROSITE; PS00092; N6\_MTASE; 1.

KW Transferase; Methyltransferase; Restriction system.

FT SEQUENCE 334 AA; 38382 MW; 8BCF1A1A35F02E89 CRC64;

SQ

Query Match 54.1%; Score 33; DB 1; Length 334;

Best Local Similarity 66.7%; Pred. No. 40;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CLATGSARQ 9

Db 142 CLAAGAAEQ 150

```

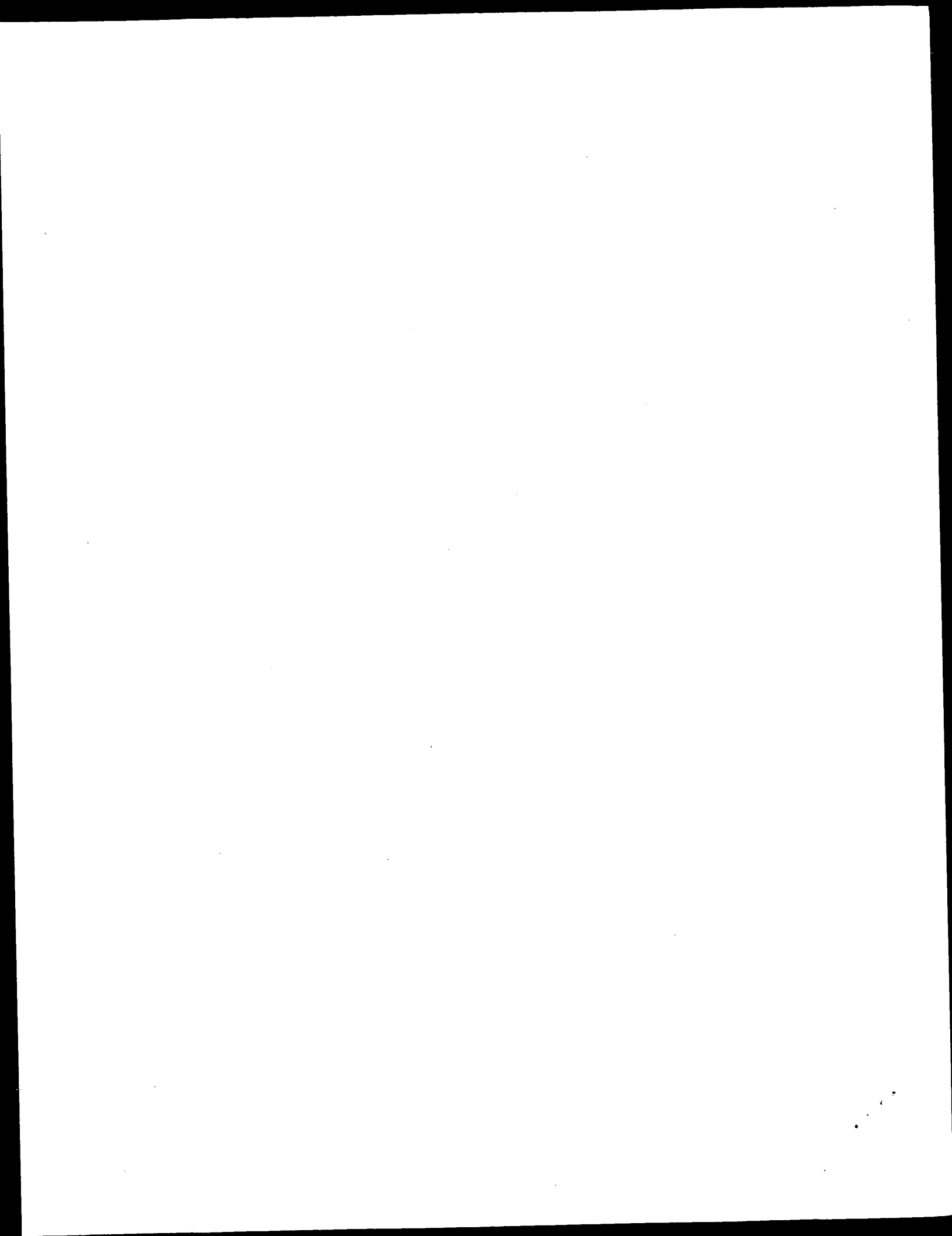
DT 01*OCT-1996 (Rel. 34, Last annotation update)
DE Putative minor fimbrial subunit pmfE precursor.
GN PMFE.
OS Proteus mirabilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Proteus.
OX NCBI_TaxID=584;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HI4320;
RX MEDLINE=95047519; PubMed=7959033;
RA Massad G., Mobley H.L.T.;
RT "Genetic organization and complete sequence of the Proteus mirabilis
  pmf fimbrial operon."
RL Gene 150:101-104(1994).
CC -|- SUBCELLULAR LOCATION: Fimbria.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; Z35428; CAA84592.1; -.
KW Fimbria; Signal.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 357 PUTATIVE MINOR FIMBRIAL SUBUNIT PMFE.
SQ SEQUENCE 357 AA; 38876 MW; 59AD7E566D4899AA CRC64;

Query Match
Best Local Similarity 54.1%; Score 33; DB 1; Length 357;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 LATGSARQLTF 12
Db :::::||||
154 ISSGSGQLTF 164

```

Search completed: June 6, 2002, 06:22:29  
Job time: 363 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:22:08 ; Search time 41.63 Seconds  
(without alignments)  
49,866 Million cell updates/sec

Title: US-08-881-509-8  
Perfect score: 61  
Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL19:\*\*

- 1: sp\_archaea:\*\*
- 2: sp\_bacteria:\*\*
- 3: sp\_fungi:\*\*
- 4: sp\_human:\*\*
- 5: sp\_invertebrate:\*\*
- 6: sp\_mammal:\*\*
- 7: sp\_mhc:\*\*
- 8: sp\_organelle:\*\*
- 9: sp\_phase:\*\*
- 10: sp\_plant:\*\*
- 11: sp\_rodent:\*\*
- 12: sp\_virus:\*\*
- 13: sp\_vertebrate:\*\*
- 14: sp\_unclassified:\*\*
- 15: sp\_virus:\*\*
- 16: sp\_bacteriaph:\*\*
- 17: sp\_archaeap:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	62.3	221	3 Q9UUQ2	Q9uuq2 penicillium
2	38	62.3	394	17 Q27058	Q27058 methanotroph
3	37	60.7	1252	11 Q9EQS9	Q9eqs9 mus musculus
4	37	60.7	1252	11 Q9JLI1	Q9jli1 mus musculus
5	37	60.7	1253	11 Q9EQS8	Q9eqs8 mus musculus
6	37	60.7	1409	2 Q52661	Q52661 escherichia
7	37	60.7	1426	2 Q93V17	Q93v17 escherichia
8	36	59.0	357	16 Q9KV40	Q9kv40 vibrio chol
9	35	58.2	1168	16 Q9YX08	Q9yx08 streptococ
10	35	57.4	138	17 Q9HHZ4	Q9hhz4 halobacteri
11	35	57.4	309	5 Q97470	Q97470 dictyosteli
12	35	57.4	423	16 Q9RWU3	Q9rwu3 deinococcus
13	35	57.4	641	4 Q95186	Q95186 homo sapien
14	35	57.4	660	3 Q9C1M9	Q9cim9 aspergillus
15	35	57.4	1028	6 Q27966	Q27966 bos taurus
16	35	57.4	1028	6 Q28138	Q28138 bos taurus

17	35	57.4	1028	11 Q63355	Q63355 rattus norv
18	35	57.4	1044	11 Q9ERB6	Q9erb6 mus musculu
19	34	55.7	84	2 Q9JAG7	Q9jag7 uncultured
20	34	55.7	91	2 Q24790	Q24790 lactobacill
21	34	55.7	91	2 Q48499	Q48499 lactobacill
22	34	55.7	208	2 Q9WWM0	Q9wwm0 synecococc
23	34	55.7	266	10 Q9AWX0	Q9awx0 oryza sativ
24	34	55.7	300	12 Q89582	Q89582 bovine herp
25	34	55.7	347	5 Q9W1A3	Q9w1a3 drosophila
26	34	55.7	354	16 Q9PH68	Q9ph68 xylella fas
27	34	55.7	356	16 Q9PH69	Q9ph69 xylella fas
28	34	55.7	396	2 Q9ZFL5	Q9zfl5 streptomyce
29	34	55.7	422	16 Q9HVS7	Q9hvs7 pseudomonas
30	34	55.7	445	10 Q9SVK1	Q9svk1 arabidopsis
31	34	55.7	480	10 Q9MSG4	Q9msg4 euphorbia e
32	34	55.7	501	10 Q9AYN6	Q9ayn6 oryza sativ
33	34	55.7	520	3 Q12610	Q12610 emericella
34	34	55.7	646	10 Q9SLD0	Q9slid0 arabidopsis
35	34	55.7	749	10 Q9LHT9	Q9lht9 oryza sativ
36	34	55.7	814	2 Q9Y129	Q9y129 arthrobacte
37	34	55.7	1145	16 Q9RQ05	Q9rzq5 deinococcus
38	34	55.7	1849	2 Q9S4K2	Q9s4k2 lactobacill
39	33.5	54.9	350	16 Q53940	Q53940 mycobacteri
40	33	54.1	77	3 Q96WD2	Q96wd2 pseudogibel
41	33	54.1	220	3 Q9HDX5	Q9hdx5 schizosacch
42	33	54.1	238	10 Q9LXD8	Q9ldx8 oryza sativ
43	33	54.1	244	2 Q9X4U4	Q9x4u4 streptococ
44	33	54.1	261	16 Q9HY27	Q9hy27 pseudomonas
45	33	54.1	270	16 Q53851	Q53851 mycobacteri

## ALIGNMENTS

RESULT 1

Q9UUQ2 ID Q9UUQ2 PRELIMINARY; PRT; 221 AA.

AC Q9UUQ2;

DT 01-MAY-2000 (T-EMBLrel. 13, Created)

DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)

DE 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)

DE XYLANASE A (EC 3.2.1.8).

GN XNA.

OS Penicillium sp. 40.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.

OX NCBI\_TaxID=110511;

RN [1]

RP SEQUENCE FROM N.A.

RA Kimura T., Sakka K., Ohmiya K.;

RT "Acidophilic xylanase A from Penicillium sp.40.";

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB035540; BAA88421.1;

DR HSSP; P36217; 1XYO.

DR InterPro; IPR001137; Glyco\_hydro\_11.

DR Pfam; PF00457; Glyco\_hydro\_11.

DR PRINTS; PR00911; GLHYDRLASE11.

DR PROSITE; PS00776; GLYCOSYL\_HYDROL\_F11\_1; 1.

KW Hydrolase; Glycosidase.

SQ SEQUENCE 221 AA; 24171 MW; 98BDD5BCEC860AD CRC64;

Query Match 62.3%; Score 38; DB 3; Length 221;  
Best Local Similarity 70.0%; Pred. No. 19;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Caps 0;

QY 3 ATGSARQLTF 12  
:|||||:|

Db 84 STGSARDITF 93

RESULT 2  
Q27058

DR SMART; SM00060; FN3; 5.  
DR SMART; SM00409; IG; 4.  
DR SMART; SM00408; IGC2; 4.  
DR SMART; SM00410; IG-like; 2.  
KW Immunoglobulin domain; Repeat.  
SQ SEQUENCE 1252 AA; 134764 MW; BA5292393483AB73 CRC64;

Query Match 60.7%; Score 37; DB 11; Length 1252;  
Best Local Similarity 77.8%; Pred. No. 1.7e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLATGSARQ 9  
DB 211 CVATNSARQ 219  
1:11 1111

RESULT 4  
Q9JL11 PRELIMINARY; PRT; 1252 AA.  
ID Q9JL11  
AC Q9JL11  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE NEIGHBOR OF PUNC E11 PROTEIN.  
GN NOPE.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RC SEQUENCE FROM N.A.  
RC STRAIN=FVB;  
RX MEDLINE=20175427; PubMed=10708514;  
RA Salbaum J.M., Kappen C.;  
RT "Cloning and expression of Nope, a new mouse gene of the  
RT immunoglobulin superfamily related to guidance receptors.";  
RL Genomics 64:15-23(2000).  
DR EMBL; AF176694; AAF65930.1; -.  
DR HSSP; P56276; 1TLK.  
DR MGD; MGI:1858497; Nope.  
DR InterPro; IPR003962; FNIII\_repeat.  
DR InterPro; IPR003961; FN\_III.  
DR InterPro; IPR003598; Ig\_c2.  
DR InterPro; IPR003600; Ig\_like.  
DR InterPro; IPR003006; Ig\_MHC.  
DR Pfam; PF00041; fn3; 5.  
DR Pfam; PF00047; Ig; 4.  
DR PRINTS; PR00014; FNTYPEIII.  
DR SMART; SM00060; FN3; 5.  
DR SMART; SM00408; IGC2; 3.  
DR SMART; SM00410; IG-like; 1.  
KW Immunoglobulin domain; Repeat.  
SQ SEQUENCE 1252 AA; 134759 MW; 11948773277B76B2 CRC64;

Query Match 60.7%; Score 37; DB 11; Length 1252;  
Best Local Similarity 77.8%; Pred. No. 1.7e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLATGSARQ 9  
DB 211 CVATNSARQ 219  
1:11 1111

RESULT 5  
Q9EQS8 PRELIMINARY; PRT; 1253 AA.  
ID Q9EQS8  
AC Q9EQS8;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE DDM36E.

ID Q27058 PRELIMINARY; PRT; 394 AA.  
AC Q27058;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE ENDO-1,4-BETA-GLUCANASE RELATED PROTEIN.  
DE MTH977.  
OS Methanothermobacter thermoautotrophicus.  
OS Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
OC Methanothermobacter.  
OX NCBI\_TaxID=145262;  
RN [1]  
RC SEQUENCE FROM N.A.  
RC STRAIN=DELTA H;  
RX MEDLINE=98037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,  
RA Aldredge T., Bashirzaden R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Nollung J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
RT deltaH: functional analysis and comparative genomics.";  
RT J. Bacteriol. 179:7135-7155(1997).  
DR EMBL; AF000871; AAB85473.1; -.  
DR Complete proteome.  
SQ SEQUENCE 394 AA; 42683 MW; B4DD13CF7CAE9B8B CRC64;

Query Match 62.3%; Score 38; DB 17; Length 394;  
Best Local Similarity 63.6%; Pred. No. 33;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 LATGSARQLTF 12  
DB 232 LATGETREVTF 242  
1111 1:11

RESULT 3  
Q9EQS9 PRELIMINARY; PRT; 1252 AA.  
ID Q9EQS9  
AC Q9EQS9;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE DDM36.  
GN NOPE OR DDM36.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RC SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6BA;  
RA Murakami H., Nakamura T., Nakayama T., Yamamoto H., Hosaka T.,  
RA Aoyama T., Nagayama S., Oka M., Kiyono T., Sasaki M.S., Nakamura T.,  
RA Toguchida J.;  
RT "Up-regulation of a ras effector and down-regulation of a cell  
RT adhesion molecule are associated with transformatation of osteoblasts.";  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB052620; BAB19278.1; -.  
DR HSSP; P56276; 1TLK.  
DR MGD; MGI:1858497; Nope.  
DR InterPro; IPR003962; FNIII\_repeat.  
DR InterPro; IPR003961; FN\_III.  
DR InterPro; IPR003599; Ig.  
DR InterPro; IPR003598; Ig\_c2.  
DR InterPro; IPR003600; Ig\_like.  
DR InterPro; IPR003006; Ig\_MHC.  
DR Pfam; PF00041; fn3; 5.  
DR Pfam; PF00047; Ig; 4.  
DR PRINTS; PR00014; FNTYPEIII.  
DR

GN NOPE OR DDM35E.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/CBA;  
 RA Murakami H., Nakamata T., Nakayama T., Yamamoto H., Hosaka T.,  
 RA Aoyama T., Nagayama S., Oka M., Kiyono T., Sasaki M.S., Nakamura T.,  
 RA Tsuchida J.;  
 RT "Up-regulation of a ras effector and down-regulation of a cell  
 adhesion molecule are associated with transformation of osteoblasts.";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB052621; BAB19279.1;  
 DR HSP; P56276; ITLK.  
 DR MGD; MGI:1858497; Nope.  
 DR InterPro: IPR003962; FNIII\_repeat.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR003599; Ig.  
 DR InterPro: IPR003598; Ig\_c2.  
 DR InterPro: IPR003600; Ig\_like.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR Pfam; PF00041; fn3; 5.  
 DR PRINTS; PR00047; Ig; 4.  
 DR PRINTS; PR00014; FNTYPEIII.  
 DR SMART; SM00060; FN3; 5.  
 DR SMART; SM00409; IG; 4.  
 DR SMART; SM00408; IGC2; 4.  
 DR SMART; SM00410; IG\_Like; 2.  
 KW Immunoglobulin domain; Repeat.  
 SQ SEQUENCE 1253 AA; 134894 MW; 2D5B254A42B9AEE CRC64;

Query Match 60.7%; Score 37; DB 11; Length 1253;  
 Best Local Similarity 77.8%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 CLATGSARQ 9  
 Db 211 CVATNSARQ 219  
 |||||  
 |||||

RESULT 6  
 ID O52661 PRELIMINARY; PRT; 1409 AA.  
 AC O52661;  
 DT 01-JUN-1998 (TRENBLrel. 06, Created)  
 DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
 DE CORE PROTEIN.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EC50;  
 RA Wang Y.D., Zhao S., Hill C.W.;  
 RT "Rhs elements comprise three subfamilies which diverged prior to  
 acquisition by Escherichia coli.";  
 RL J. Bacteriol. 180:4102-4110(1998).  
 DR EMBL; AF044499; AAC32464.1;  
 DR InterPro: IPR001826; Rhs.  
 DR PRINTS; PR00394; RHSPROTEIN.  
 SQ SEQUENCE 1409 AA; 158612 MW; 2D5D82E8A11BDEA8 CRC64;

Query Match 60.7%; Score 37; DB 2; Length 1409;  
 Best Local Similarity 70.0%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 ATGSARQLTF 12  
 Db 531 ATGSTQMTW 540  
 |||||

RESULT 7  
 ID O93V17 PRELIMINARY; PRT; 1426 AA.  
 AC O93V17;  
 DT 01-DEC-2001 (TRENBLrel. 19, Created)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
 DE CORE PROTEIN.  
 GN RUSD.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K-12;  
 RX MEDLINE=89123133; PubMed=2644231;  
 RA Sadosky A.B., Davidson A., Lin R.J., Hill C.W.;  
 RT "rhs gene family of Escherichia coli K-12.";  
 RL J. Bacteriol. 171:636-642(1989).  
 DR EMBL; L19084; AAA24544.1;  
 SQ SEQUENCE 1426 AA; 159724 MW; 4F6D7084470FBC2D CRC64;

Query Match 60.7%; Score 37; DB 2; Length 1426;  
 Best Local Similarity 70.0%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 ATGSARQLTF 12  
 Db 528 ATGSTQMTW 537  
 |||||

RESULT 8  
 ID O9KV40 PRELIMINARY; PRT; 357 AA.  
 AC O9KV40;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
 DE UDP-N-ACETYLENOLPYRUVOYLGLUCOSAMINE REDUCTASE.  
 GN VC0318.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 cholerae.";  
 RL Nature 406:477-483(2000).  
 DR EMBL; AE004120; AAF93491.1;  
 DR HSP; P08373; 2MBR.

	TIGR; VC0318; -; InterPro: IPR003170; MurB. Pfam: PF02215; MurB; 1. Complete proteome. SEQUENCE 357 AA; 39351 MW; C5DC7E1EEB934E0C CRC64;	Query Match Best Local Similarity 59.0%; Score 36; DB 16; Length 357; Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY	1 CLATGSARQLT 11      :    154 CLETGTVKRLT 164	
Db		
RESULT 9		
Q99YQ8 ID	PRELIMINARY; PRT; 1168 AA.	
AC Q99YQ8;		
DT 01-JUN-2001 (TREMBLrel. 17, Created)		
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE PUTATIVE BETA-GALACTOSIDASE.		
DN SPY1586		
OS Streptococcus pyogenes.		
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OC Streptococcus.		
OX NCBI_TaxID=1314;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;		
RX MEDLINE=21192684; PubMed=11296296;		
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Lyon K.,		
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,		
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,		
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;		
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";		
RT Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).		
DR EMBL: AE006590; AAK34368.1;		
DR InterPro: IPR001649; Glyco_hydro_2.		
DR Pfam: PF00703; Glyco_hydro_2; 1.		
DR PRINTS: PR00132; GLHYDLASE2.		
KW Complete proteome.		
SQ SEQUENCE 1168 AA; 131946 MW; F323127F2449CEA3 CRC64;		
Query Match	58.2%; Score 35.5; DB 16; Length 1168;	
Best Local Similarity 75.0%; Pred. No. 3e+02; 2; Indels 1; Gaps 1;		
Matches 9; Conservative 0; Mismatches 0;		
QY 1 CLATGSARQLTF 12 		
Db 318 CLETG-FRLQTF 328		
RESULT 10		
Q9HHZ4 ID	PRELIMINARY; PRT; 138 AA.	
AC Q9HHZ4;		
DT 01-MAR-2001 (TREMBLrel. 16, Created)		
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)		
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)		
DE VNG6157H.		
GN VNG6157H.		
OS Halobacterium sp. (strain NRC-1).		
OG plasmid pNRC200.		
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;		
OC Halobacterium		
OX NCBI_TaxID=64091;		
RN [1]		
RP SEQUENCE FROM N.A.		
RX MEDLINE=20504483; PubMed=11016950;		

RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M., Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J., Swartzell S., Weir D., Hall J., Dahl T.A., Welte R., Goo Y.A., Leihhauser B., Keller K., Cruz R., Danson M.J., Hough D.W., Madenbach D.G., Jablonksi P.E., Krebs M.P., Angevine C.M., Dale H., Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H., Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D., Eberhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.; RT "Genomic sequence of Halobacterium species NRC-1."; Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000); DR EMBL: AE005152; AAG20828.1; -. KW Plasmid; Complete proteome. SQ SEQUENCE 138 AA; 14950 MW; 3E347A1BB44F2DAC CRC64;	Query Match Best Local Similarity 57.4%; Score 35; DB 17; Length 138; Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;	
QY 1 CLATGSARQLT 11   :		
Db 36 CLSQGEARSIT 46		
RESULT 11		
O97470 PRELIMINARY; PRT; 309 AA.		
ID O97470;		
DT 01-MAY-1999 (TREMBLrel. 10, Created)		
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)		
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)		
DE ADP/ATP TRANSLOCASE.		
OS Dictyostelium discoideum (Slime mold).		
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.		
OX NCBI_TaxID=44689;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN=AX2;		
RX MEDLINE=95192826; PubMed=10092866;		
RA Bof M., Brandolin G., Satre M., Klein G.;		
RA "The mitochondrial adenine nucleotide translocator from Dictyostelium"		
RT discoidium. Functional characterization and DNA sequencing.";		
RL Eur. J. Biochem. 259:795-800(1999).		
CC -! SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).		
CC -! SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.		
CC EMBL: AF100676; AAC79081.1; -.		
DR EMBL: AF039211; AAC77879.1; -.		
DR InterPro: IPR001993; Mitoch_carrier.		
DR InterPro: IPR002067; Mit_carrier.		
DR Pfam: PF00153; mito_carr; 3.		
DR PRINTS: PR00926; MITOCARRIER.		
DR PROSITE: PS00215; MITOCH_CARRIER; 3.		
KW Inner membrane; Mitochondrion; Repeat; Transmembrane; Transport.		
SQ SEQUENCE 309 AA; 33468 MW; D623DDBD4BA49474 CRC64;		
Query Match	57.4%; Score 35; DB 5; Length 309;	
Best Local Similarity 70.0%; Pred. No. 1e+02;		
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;		
QY 2 LATGSARQLT 11 :		
Db 148 VGTSARQFT 157		
RESULT 12		
O9RWU3 ID	PRELIMINARY; PRT; 423 AA.	
ID O9RWU3		
AC O9RWU3;		
DT 01-MAY-2000 (TREMBLrel. 13, Created)		
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)		

DE SPORULATION PROTEIN SPOIID-RELATED PROTEIN.  
 GN DR0572.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
 OX NCBI\_TaxID=1299;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R1.  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Ramphile W., Crosby M., Shen M.,  
 RA Vamatheva J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
 RA Makarewicz K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 RT radiodurans R1.";  
 RL Science 286:1571-1577(1999).  
 DR EMBL; AE001915; AAF10150.1;  
 DR TIGR; DR0572;  
 KW Complete proteome.  
 SQ SEQUENCE 423 AA; 43100 MW; D42323D17EFOE081 CRC64;

Query Match 57.4%; Score 35; DB 16; Length 423;  
 Best Local Similarity 70.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Qy 3 ATGSARQLTF 12  
 I I I I I I I I  
 Db 305 ASGRAQLTF 314

RESULT 13  
 Q95186 PRELIMINARY; PRT; 641 AA.  
 AC Q95186;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE FATY ACID TRANSPORT PROTEIN.  
 GN FATP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HEART;  
 RX MEDLINE=99096471; PubMed=9878842;  
 RA Fitcher B.A., Riedel H.D., Young K.C., Stremmel W.;  
 RT "Tissue distribution and cDNA cloning of a human fatty acid transport  
 RT protein (hsFATP4).";  
 RL Biochim. Biophys. Acta 1443:381-385(1998).  
 DR EMBL; AF055899; AAD11623.1;  
 DR InterPro; IPR000873; AMP-bind.  
 DR InterPro; IPR000566; Lipocln\_cytFABP.  
 DR Pfam; PF00501; AMP-binding; 2.  
 DR PROSITE; PS00455; AMP-BINDING; 1.  
 DR PROSITE; PS00213; LIPOCALIN; UNKNOWN; 1.  
 SQ SEQUENCE 641 AA; 71431 MW; 8C24F76C9BF81378 CRC64;

Query Match 57.4%; Score 35; DB 4; Length 641;  
 Best Local Similarity 58.3%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy 1 CLATGSARQLTF 12  
 I I I I I I I I  
 Db 168 CLTTSARALVF 179

RESULT 14  
 Q9C1M9 PRELIMINARY; PRT; 660 AA.  
 AC Q9C1M9;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE ALPHA-L-RHAMNOSIDASE A PRECURSOR (EC 3.2.1.40).  
 GN RHAA.  
 OS Aspergillus aculeatus.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_TaxID=5053;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21218536; PubMed=11319105;  
 RA Manzanarez P., van Den Broeck H.C., de Graaff L.H., Visser J.;  
 RT "Purification and Characterization of Two Different alpha-L-  
 RT Rhamnosidases, RhaA and RhaB, from Aspergillus aculeatus.";  
 RL Appl. Environ. Microbiol. 67:2230-2234(2001).  
 DR EMBL; AF284761; AAK16249.1;  
 KW Signal; Hydrolase; Glycosidase.  
 FT SIGNAL 1 19 POTENTIAL.  
 SQ SEQUENCE 660 AA; 71218 MW; A550A9F4D0E42984 CRC64;

Query Match 57.4%; Score 35; DB 3; Length 660;  
 Best Local Similarity 70.0%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Qy 3 ATGSARQLTF 12  
 I I I I I I I I  
 Db 604 ATGTFQQLTF 613

RESULT 15  
 Q27966 PRELIMINARY; PRT; 1028 AA.  
 AC Q27966;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE MYOSIN I.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=ADRENAL GLAND;  
 RX MEDLINE=94148088; PubMed=8313976;  
 RA Zhu T., Ikebe M.;  
 RT "Cloning of myosin I from bovine adrenal gland.";  
 RL FEBS Lett. 339:31-36(1994).  
 DR EMBL; U03420; AAL17565.1;  
 DR HSSP; P08799; IMND.  
 DR InterPro; IPR000048; IQ.  
 DR InterPro; IPR001609; myosin\_head.  
 DR Pfam; PF00612; IQ; 2.  
 DR Pfam; PF00063; myosin\_head; 1.  
 DR PRINTS; PR00193; MYOSINHEAVY.  
 DR PRODOM; PD000355; myosin\_head; 1.  
 DR SMART; SM00015; IQ; 2.  
 DR SMART; SM00242; MYSC; 1.  
 SQ SEQUENCE 1028 AA; 118020 MW; 5CFFEE28CF2A52D7 CRC64;

Query Match 57.4%; Score 35; DB 6; Length 1028;  
 Best Local Similarity 58.3%; Pred. No. 3.3e+02;  
 Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy 1 CLATGSARQLTF 12

us-08-881-509-8.open.rspt

Thu Jun 6 10:09:42 2002

Db 463 CLRPGATDLTF 474

Search completed: June 6, 2002, 06:22:11  
Job time: 380 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:17:20 ; Search time 52.04 seconds  
(without alignments)  
25.613 Million cell updates/sec

Title: US-08-881-509-8  
Perfect score: 61  
Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	100.0	12	19	AAW47591
2	50	82.0	267	19	AAW47588
3	42.5	69.7	13	19	AAW47590
4	41	67.2	153	17	AAW47586
5	41	67.2	263	17	AAW47585
6	39	63.9	179	21	AAW47584
7	37.5	61.5	215	22	ABG11979
8	37	60.7	119	22	ABG23742
9	37	60.7	740	22	AAW47591
10	37	60.7	932	22	AAW47591
11	37	60.7	1250	22	AAW47591

12	37	60.7	1252	22	AAW47591
13	37	60.7	1395	22	ABG23742
14	37	60.7	1426	21	AAW47591
15	36	59.0	108	21	AAW47591
16	36	59.0	217	22	AAW47591
17	36	59.0	391	22	AAW47591
18	36	59.0	391	22	AAW47591
19	36	59.0	391	22	AAW47591
20	36	59.0	391	22	AAW47591
21	36	59.0	391	22	AAW47591
22	36	59.0	406	22	ABG19026
23	35	57.4	428	18	AAW16543
24	35	57.4	23	22	ABG31905
25	35	57.4	23	22	ABG31905
26	35	57.4	23	22	ABG31905
27	35	57.4	23	22	ABG31905
28	35	57.4	23	22	ABG31905
29	35	57.4	23	22	ABG31905
30	35	57.4	23	22	ABG31905
31	35	57.4	23	22	ABG31905
32	35	57.4	23	22	ABG31905
33	35	57.4	23	22	ABG31905
34	35	57.4	23	22	ABG31905
35	35	57.4	23	22	ABG31905
36	35	57.4	23	22	ABG31905
37	35	57.4	23	22	ABG31905
38	35	57.4	23	22	ABG31905
39	35	57.4	23	22	ABG31905
40	35	57.4	23	22	ABG31905
41	35	57.4	23	22	ABG31905
42	35	57.4	23	22	ABG31905
43	35	57.4	23	22	ABG31905
44	34	55.7	61	22	AAW46008
45	34	55.7	170	21	AAW99741

#### ALIGNMENTS

RESULT 1

AAW47591  
ID AAW47591 standard; peptide; 12 AA.

AC AAW47591;

DT 26-JUN-1998 (first entry)

XX T-cell receptor CDR3 alpha-region.

DE Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;

KW Prevention; therapy; tumour disease; renal cell carcinoma;

XX CDR3.

OS Homo sapiens.

XX DE19625191-A1.

XX 02-JAN-1998.

PF 24-JUN-1996; 96DE-1025191.

XX 24-JUN-1996; 96DE-1025191.

XX (BOEF) BOEHRINGER MANNHEIM GMBH.

XX Schendel D;

XX WPI; 1998-053442/06.

DR N-PSDB; AAV18708.

XX Human T-cell receptor nucleic acids and poly:peptide(s) - for

PT diagnosis or therapy, especially of renal cell carcinoma

XX

Mouse Nope (neighb  
Novel human diagno  
E. coli proliferat  
Pinus radiata tran  
Penicillium griseo  
C. glutamicum meta  
C. glutamicum prote  
Corynebacterium gl  
Corynebacterium gl  
50K-cellulase from  
Novel human diagno  
Peptide #4556 enco  
Peptide #4649 enco  
Protein #4455 enco  
Human bone marrow  
Peptide #4542 enco  
Peptide #4657 enco  
Peptide #4425 enco  
Human membrane tra  
Human protein sequ  
Human ORF2520  
Human FATP4 SEQ ID  
Human FATP4 SEQ ID  
Amino acid sequenc  
Human FATP4 SEQ ID  
Human FATP4 SEQ ID  
Human FATP4 SEQ ID  
Novel human diagno  
Human poly:peptide  
Protonibacterium  
Corn PI metabolism

PS Example 1; Page 17; 30pp; German.

XX The present sequence is the CDR3 alpha-region of a human  
CC T-cell receptor (TCR), which can be used in the diagnosis,  
CC monitoring, prevention and therapy of a tumour disease,  
CC specifically renal cell carcinoma.

XX Sequence 12 AA;

Query Match 100.0%; Score 61; DB 19; Length 12;  
Best Local Similarity 100.0%; Pred. No. 4.8e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLATGSARQLTF 12  
| | | | | | | | | | | | | |  
Db 1 clatgsarqltf 12

RESULT 2  
AAW47588  
ID AAW47588 standard; Protein; 267 AA.

XX AAW47588;  
AC  
XX 26-JUN-1998 (first entry)  
DT  
XX T-cell receptor alpha-chain.  
DE

XX Alpha-chain; human; T-cell receptor; TCR; diagnosis; monitoring;  
KW prevention; therapy; tumour disease; renal cell carcinoma.  
XX

OS Homo sapiens.  
XX  
XX DEL19625191-A1.  
FN  
XX 02-JAN-1998.  
PD

XX 24-JUN-1996; 96DE-1025191.  
XX  
XX 24-JUN-1996; 96DE-1025191.  
PR

XX (BOEF) BOEHRINGER MANNHEIM GMBH.  
PA

XX Schendel D;  
PI

XX WPI; 1998-053442/06.  
DR  
XX N-PSDB; AAV18705.  
DR

XX Human T-cell receptor nucleic acids and poly:peptide(s) - for  
PT diagnosis or therapy, especially of renal cell carcinoma

XX Example 1; Pages 11-13; 30pp; German.  
PS

XX The present sequence is the alpha-chain of a human T-cell  
CC receptor (TCR), which can be used in the diagnosis, monitoring,  
CC prevention and therapy of a tumour disease, specifically renal  
CC cell carcinoma.  
XX

XX Sequence 267 AA;

Query Match 82.0%; Score 50; DB 19; Length 267;  
Best Local Similarity 83.3%; Pred. No. 0.17;  
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CLATGSARQLTF 12  
| | | | | | | | | | | | | |  
Db 105 clvgsarqltf 116

RESULT 3  
AAW47590

ID AAW47590 standard; peptide; 13 AA.

XX AAW47590;

XX 26-JUN-1998 (first entry)  
DT

XX T-cell receptor CDR3 alpha-region.  
DE

XX Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;  
KW prevention; therapy; tumour disease; renal cell carcinoma;  
KW CDR3.

XX Homo sapiens.  
OS

XX DEL19625191-A1.  
PN

XX 02-JAN-1998.  
PD

XX 24-JUN-1996; 96DE-1025191.  
PF

XX 24-JUN-1996; 96DE-1025191.  
PR

XX (BOEF) BOEHRINGER MANNHEIM GMBH.  
PA

XX Schendel D;  
PI

XX WPI; 1998-053442/06.  
DR  
XX N-PSDB; AAV18707.  
DR

XX Human T-cell receptor nucleic acids and poly:peptide(s) - for  
PT diagnosis or therapy, especially of renal cell carcinoma

XX Example 1; Page 17; 30pp; German.  
PS

XX The present sequence is the CDR3 alpha-region of a human  
CC T-cell receptor (TCR), which can be used in the diagnosis,  
CC monitoring, prevention and therapy of a tumour disease,  
CC specifically renal cell carcinoma.  
XX

XX Sequence 13 AA;

Query Match 69.7%; Score 42.5; DB 19; Length 13;  
Best Local Similarity 76.9%; Pred. No. 0.18;  
Matches 10; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 1 CLATGSARQLTF 12  
| | | | | | | | | | | | | |  
Db 1 clvgsarqltf 13

RESULT 4  
AAW47590  
ID AAW47590 standard; Protein; 153 AA.

XX AAW47590;

XX 13-JAN-1997 (first entry)  
DT

XX Alpha chain of soluble secreted D10 dual chain T-cell receptor.  
DE

XX T-cell receptor; TCR; pathogenic T-cell; killer T-cell;  
KW T-lymphocyte; inhibition; activation; graft rejection; antigen;  
KW diabetes; assay; detection; heterodimer; alpha chain; beta chain.  
XX

XX Mus musculus.  
OS

XX Key Location/Qualifiers  
XX Region 147  
XX /label= C-alpha

XX WO9621028-A2.



PD 11-JUL-1996.  
 XX  
 PF 28-DEC-1995; 95WO-US16937.  
 XX  
 PR 03-JAN-1995; 95US-0367589.  
 XX  
 PA (PROC-) PROCEPT INC.  
 XX  
 PI Banerji J, Brauer P, Jesson M, Jones B, Khandekar S;  
 PI McKeever U, Naylor J;  
 XX  
 DR WPI; 1996-334003/33.  
 DR N-PSDB; AAT31598.  
 XX  
 PT Soluble heterodimeric T cell receptor contg. linked alpha and beta  
 PT subunit - useful to reduce pathogenic T cell activation, and to  
 PT immunise mammals, e.g. to prevent graft rejection  
 XX  
 PS Claim 1; Figure 1; 120pp; English.  
 XX  
 CC A soluble heterodimeric T-cell receptor (TCR) which contains an  
 CC alpha subunit connected by disulphide bonding to a beta subunit can  
 CC be used to detect and analyse the peptide and MHC/HLA molecular  
 CC constituents of TCR ligands. The TCR is useful to reduce the  
 CC activation of pathogenic T cells in a mammal, and to immunise  
 CC against TCR antigenic structures on the surface of such cells e.g. to  
 CC prevent graft rejection. It can also be used to deplete pathogenic  
 CC T-cells or inhibit their activation in a female's offspring by  
 CC administration to the female during gestation or before weaning.  
 CC The TCR and antibodies directed against it can also be used to  
 CC deplete peripheral lymphocyte T-cells involved in T-cell mediated  
 CC disease in a mammal or its offspring, particularly diabetes. The  
 CC heterodimeric TCR can also be used in an assay to detect pathogenic  
 CC T-cells.  
 XX  
 SQ Sequence 153 AA;  
 Query Match 67.2%; Score 41; DB 17; Length 153;  
 Best Local Similarity 66.7%; Pred. No. 4.8;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 CLATGSARQLTF 12  
 DB 118 caatgsfnkltf 129  
 RESULT 5  
 AAR97725  
 ID AAR97725 standard; Protein; 263 AA.  
 XX  
 AC AAR97725;  
 XX  
 DT 18-SEP-1996 (first entry)  
 XX  
 DE D10 single chain T-cell receptor.  
 XX  
 KW Single chain T-cell receptor; TCR; maltose binding protein;  
 KW MBP-D10 scTCR; fusion protein; V-alpha; V-beta; antibody; vaccine;  
 KW conalbumin.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Cleavage-site 6..7  
 FT /note= "thrombin cleavage site"  
 FT Domain 7..118  
 FT /label= V-beta  
 FT /note= "residue 8 (Ala in the wild-type) is  
 FT substituted by Ser to facilitate  
 FT thrombin cleavage"  
 FT Region 119..145  
 FT /label= Linker

FT Domain 146..257  
 FT /label= V-alpha  
 FT 258..263  
 FT /label= Hexahistidine\_tail  
 XX  
 PN WO9613593-A2.  
 XX  
 PD 09-MAY-1996.  
 XX  
 PF 26-OCT-1995; 95WO-US13770.  
 XX  
 PR 06-JUN-1995; 95US-0458131.  
 PR 26-OCT-1994; 94US-0329310.  
 PR 01-DEC-1994; 94US-0347893.  
 XX  
 PA (PROC-) PROCEPT INC.  
 XX  
 PI Banerji J, Bettencourt B, Dwyer D, Jesson M, Jones B;  
 PI Khandekar S, McKeever U, Naylor J;  
 XX  
 DR WPI; 1996-239502/24.  
 DR N-PSDB; AAT29757.  
 XX  
 PT New fusion proteins comprising a single chain T-cell receptor -  
 PT used to develop prods. for use in detection, diagnosis, functional  
 PT studies and therapy involving immune responses  
 XX  
 PS Example 1; Page 61-62; 104pp; English.  
 XX  
 CC A soluble single chain T-cell receptor (sCTCR) (AAR97725) comprises  
 CC the V-beta region of conalbumin-specific D10 T-cell receptor (ATCC  
 CC TIB 224) joined to the D10 V-alpha region via a linker peptide. It  
 CC is obtd. by PCR amplification of V-beta and V-alpha sequences, and  
 CC cloning into a vector that encodes the linker. The entire sCTCR  
 CC sequence (AAT29757) is then cloned into vector pPR998 which encodes  
 CC maltose binding protein (MBP). The MBP-D10 sCTCR fusion protein  
 CC is expressed in E. coli and purified by affinity chromatography.  
 CC Thrombin cleavage yields a soluble sCTCR useful in detection,  
 CC diagnosis, functional studies and therapy involving immune responses.  
 XX  
 SQ Sequence 263 AA;  
 Query Match 67.2%; Score 41; DB 17; Length 263;  
 Best Local Similarity 66.7%; Pred. No. 8.6;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 CLATGSARQLTF 12  
 DB 235 caatgsfnkltf 246  
 RESULT 6  
 AAR25390  
 ID AAR25390 standard; Protein; 179 AA.  
 XX  
 AC AAR25390;  
 XX  
 DT 27-NOV-2000 (first entry)  
 XX  
 DE Pinus radiata cell signalling involved protein SEQ ID NO:709.  
 XX  
 KW Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;  
 KW plant cell signalling; modulation; transgenic plant; pathogen; growth;  
 KW environmental change; development; cell proliferation; differentiation;  
 KW elongation; survival; disease resistance; nutrient metabolism.  
 XX  
 OS Pinus radiata.  
 XX  
 PN WO200042171-A1.  
 XX  
 PD 20-JUL-2000.

PF 11-JAN-2000; 2000WO-US00724.  
 XX 12-JAN-1999; 99US-0228986.  
 PR 01-NOV-1999; 99US-0162866.  
 XX (GENE-) GENESIS RES & DEV CORP LTD.  
 PA Strabala TJ, Nieuwenhuizen NJ;  
 XX WPI; 2000-476052/41.  
 XX  
 XX Isolated polynucleotide encoding a polypeptide involved in cell  
 PT signaling used for generating transgenic plants with modified responses  
 PT to external signals -  
 XX  
 XX Claim 3; Page 326; 527pp; English.  
 XX  
 XX AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide  
 CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or  
 CC pine (Pinus radiata also known as Monterey pine). The protein sequences  
 CC are involved in cell signalling. The polynucleotide and protein  
 CC sequences can be used to modify the response of plant cells to external  
 CC signals e.g. environmental changes or pathogens during the growth and  
 CC development of a plant. They can be used to modify cell proliferation,  
 CC differentiation, elongation and survival, resistance to disease and  
 CC nutrient metabolism. Examples of modifications which can be produced are  
 CC altered fruit ripening and senescence of leaves and flowers e.g. to  
 CC delay senescence and prolong the life of cut flowers or enhance  
 CC senescence of reproductive organs to engineer sterile plants. Other  
 CC modifications can be used to delay senescence in selected cell types or  
 CC organs providing fruit and vegetables which have a longer shelf life  
 CC between harvest and consumption, or to decrease branching frequency in  
 CC forest tree species giving long stretches of valuable knot-free clear  
 CC wood which can be used in solid timber furniture and veneers.  
 XX  
 XX Sequence 179 AA;  
 SQ

Query Match 63.9%; Score 39; DB 21; Length 179;  
 Best Local Similarity 58.3%; Pred. No. 14;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CLATGSARQLTF 12  
 1:1 ||||| 1:  
 Db 156 cialgsarglsy 167

RESULT 7  
 ABG11979  
 ID ABG11979 standard; Protein: 215 AA.  
 XX  
 XX ABG11979;  
 AC  
 XX 18-FEB-2002 (first entry)  
 DT  
 XX Novel human diagnostic protein #11970.  
 DE  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200175067-A2.  
 PN  
 XX 11-OCT-2001.  
 PD  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX  
 XX 31-MAR-2000; 2000US-0540217.  
 XX  
 XX 23-AUG-2000; 2000US-0649167.  
 PR  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Drmanac RT, Liu C, Tang YT;  
 PI

PI Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS76166.  
 XX  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 XX Claim 20; SEQ ID No 42338; 103pp; English.  
 XX  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: the sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 215 AA;  
 SQ

Query Match 61.5%; Score 37.5; DB 22; Length 215;  
 Best Local Similarity 69.2%; Pred. No. 32;  
 Matches 9; Conservative 2; Mismatches 1; Indels 1; Gaps 1;  
 QY 1 CLATG-SARQLTF 12  
 1:1 | 1:1 | 1:1 |  
 Db 81 clasgeslrkltf 93

RESULT 8  
 ABG23742  
 ID ABG23742 standard; Protein: 119 AA.  
 XX  
 XX ABG23742;  
 AC  
 XX 18-FEB-2002 (first entry)  
 DT  
 XX Novel human diagnostic protein #23733.  
 DE  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200175067-A2.  
 PN  
 XX 11-OCT-2001.  
 PD  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX  
 XX 31-MAR-2000; 2000US-0540217.  
 XX  
 XX 23-AUG-2000; 2000US-0649167.  
 PR  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Drmanac RT, Liu C, Tang YT;  
 PI

XX WPI: 2001-639362/73.  
 DR N-PSDB: AAS87929.  
 XX  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 XX Claim 20: SEQ ID No 54101; 103pp; English.  
 XX  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ARG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 119 AA;

Query Match 60.7%; Score 37; DB 22; Length 119;  
 Best Local Similarity 63.6%; Pred. No. 21;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 CLATGSARQLT 11  
 II IIII :I  
 Db 52 clmtgsaenvt 62  
 RESULT 9  
 AAG65913  
 ID AAG65913 standard; protein; 740 AA.  
 XX  
 AC AAG65913;  
 XX  
 DT 11-FEB-2002 (first entry)  
 XX  
 DE Amino acid sequence of GSK gene Id 27142.  
 XX  
 KW Peptide hormone; antidiabetic; anorectic; antianorectic; antiasthmatic;  
 KW antidepressant; nootropic; neuroprotectant; hypotensive; hypertensive;  
 KW cytostatic; cerebroprotective; vasotropic; human.  
 OS Homo sapiens.  
 XX  
 PN WO200172961-A2.  
 XX  
 PD 04-OCT-2001.  
 XX  
 PF 22-MAR-2001; 2001WO-US09226.  
 XX  
 PR 24-MAR-2000; 2000US-192158P.  
 PR 28-MAR-2000; 2000US-192668P.  
 PR 27-APR-2000; 2000US-200166P.  
 XX  
 XX (SMK ) SMITHKLINE BEECHAM CORP.  
 PA (SMK ) SMITHKLINE BEECHAM PLC.

XX  
 PI Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z, Kabnick KS;  
 PI Lai Y;  
 XX  
 DR WPI: 2001-639223/73.  
 DR N-PSDB: AAI67203.  
 XX  
 XX Isolated polypeptides, which may be peptide hormones, which are  
 PT identified by high throughput genome-based biology which identifies  
 PT genes and gene products as therapeutic targets for treatment of  
 PT diseases such as diabetes and cancer -  
 XX  
 PS Claim 1: Page 89-91; 99pp; English.  
 XX  
 CC The invention provides polypeptides (AAG65886-65918) which may be peptide  
 CC hormones (including insulin, growth hormones, chemokines, cytokines,  
 CC neurotrophins, integrins, kallikreins, lamins, melanins, natriuretic  
 CC hormones, neurotrophins, pituitary hormones, pleiotrophins, prostaglandins,  
 CC secretogranins, selectins, thromboglobulins, thymosins) identified by  
 CC high throughput genome-based biology and polynucleotides (AAI67176-67208)  
 CC encoding them. The polypeptides can be expressed by standard recombinant  
 CC methodology. The polypeptides are useful in the treatment of disease such  
 CC as diabetes, breast-, prostate-, colon cancer and other malignant tumors,  
 CC hyper- and hypotension, obesity, bulimia, anorexia, growth abnormalities,  
 CC asthma, manic depression, dementia, delirium, mental retardation,  
 CC Huntington's disease, Tourette's syndrome, schizophrenia, growth, mental  
 CC or sexual development disorders, and dysfunctions of the blood cascade  
 CC system including those leading to stroke. The polynucleotides may be used  
 CC as diagnostic reagents through detecting mutations in the associated gene  
 CC and for chromosome localization and for tissue expression studies. The  
 CC polypeptides and polynucleotides may also be used as vaccines.  
 XX  
 SQ Sequence 740 AA;

Query Match 60.7%; Score 37; DB 22; Length 740;  
 Best Local Similarity 77.8%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CLATGSARQ 9  
 I:II IIII  
 Db 214 cvatnsarq 222  
 RESULT 10  
 AAE05252  
 ID AAE05252 standard; Protein; 932 AA.  
 XX  
 AC AAE05252;  
 XX  
 DT 12-SEP-2001 (first entry)  
 XX  
 DE Mouse Nope (neighbour of punc ell) extracellular domain.  
 XX  
 KW Mouse: Nope; neighbour of punc ell; cytostatic; neuroprotective; vaccine;  
 KW gene therapy; cerebroprotective; colonic cancer; mental retardation;  
 KW tumour suppressor; chromosome 9; transgenic animal; genetic disorder;  
 KW obesity; Bardet-Biedl syndrome; autosomal recessive disorder; retinitis;  
 KW polydactyly; pigmentosa; hypogonadism; hypertension; diabetes mellitus;  
 KW renal anomaly; cardiovascular anomaly; extracellular domain.  
 XX  
 OS Mus musculus.  
 XX  
 PN WO200149714-A2.  
 XX  
 PD 12-JUL-2001.  
 XX  
 PF 26-OCT-2000; 2000WO-US29698.  
 XX  
 PR 04-JAN-2000; 2000US-0174496.  
 PR 19-MAY-2000; 2000US-0205789.  
 XX  
 PA (NEUR-) NEUROSCIENCES RES FOUND INC.

XX Salbaum JM;  
 PI WPI; 2001-441846/47.  
 XX N-PSDB; AAD10022.  
 DR Murine Nope polypeptides and nucleic acids useful for preventing,  
 PT diagnosing and treating colonic cancer and Bardet-Biedl syndrome -  
 PS Claim 2; Page 81-83; 99pp; English.  
 XX The present invention relates to Nope (neighbour of punc ell) which is  
 CC used in the prevention, treatment and diagnosis of diseases associated  
 CC with inappropriate Nope expression such as cancers especially colonic  
 CC cancer and genetic disorders, as Nope is thought to be a tumour  
 CC suppressor. Nope gene is located on chromosome 9 and is used in gene  
 CC therapy. Nope is used as vaccine. Nope gene may be administered to treat  
 CC diseases by rectifying mutations or deletions in a patient's genome that  
 CC affect the activity of Nope by expressing inactive proteins or to  
 CC supplement the patients own production of Nope polypeptides. Nope gene  
 CC is used to study the expression and function of Nope polypeptides and  
 CC their role in metabolism through the creation of transgenic animal  
 CC models. The anti-Nope antibodies and Nope antagonists may also be used  
 CC to down regulate Nope expression and activity for the treatment of  
 CC Bardet-Biedl syndrome which is an autosomal recessive disorder  
 CC characterised by mental retardation, obesity, polydactyly, retinitis  
 CC pigmentosa and hypogonadism. Patients with Bardet-Biedl syndrome have a  
 CC high incidence of hypertension, diabetes mellitus and renal and  
 CC cardiovascular anomalies. The present sequence is mouse Nope (neighbour  
 CC of punc ell) extracellular domain.  
 XX Sequence 932 AA;  
 SQ

Query Match 60.7%; Score 37; DB 22; Length 932;  
 Best Local Similarity 77.8%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLATGSARQ 9  
 I:| | | | |  
 Db 190 cvatnsarq 198

RESULT 11  
 AAG65914  
 ID AAG65914 standard; protein; 1250 AA.  
 XX AC AAG65914;  
 XX 11-FEB-2002 (first entry)  
 XX DE Amino acid sequence of GSK gene Id 27142.  
 XX KW Peptide hormone; antidiabetic; anorectic; antianorectic; antiasthmatic;  
 XX KW antipressant; nootropic; neuroprotectant; hypotensive; hypertensive;  
 KW cytotatic; cerebroprotective; vasotropic; human.  
 XX OS Homo sapiens.  
 XX WO200172961-A2.  
 XX PD 04-OCT-2001.  
 XX PF 22-MAR-2001; 2001WO-US09226.  
 XX PR 24-MAR-2000; 2000US-192158P.  
 XX PR 28-MAR-2000; 2000US-192668P.  
 XX PR 27-APR-2000; 2000US-200166P.  
 XX {SMIK } SMITHKLINE BEECHAM CORP.  
 PA {SMIK } SMITHKLINE BEECHAM PLC.  
 XX Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z, Kabnick KS;  
 PI

Lai Y;

WPI; 2001-639223/73.

N-PSDB; AAI67204.

Isolated polypeptides, which may be peptide hormones, which are  
 identified by high throughput genome-based biology which identifies  
 genes and gene products as therapeutic targets for treatment of  
 diseases such as diabetes and cancer -

Claim 1; Page 91-94; 99pp; English.

The invention provides polypeptides (AAG65886-65918) which may be peptide  
 hormones (including insulin, growth hormones, chemokines, cytokines,  
 neuropeptides, integrins, kallikreins, lamins, melanins, natruiretic  
 hormones, neuropepsin, pituitary hormones, pleiotrophins, prostaglandins,  
 secretogranins, selectins, thromboglobulins, thymosins) identified by  
 high throughput genome-based biology and polynucleotides (AAI67176-67208)  
 encoding them. The polypeptides can be expressed by standard recombinant  
 methodology. The polypeptides are useful in the treatment of disease such  
 as diabetes, breast-, prostate-, obesity, colon cancer and other malignant tumors,  
 hyper- and hypotension, dementia, delirium, mental retardation, mental  
 asthma, manic depression, Tourette's syndrome, schizophrenia, growth, mental  
 Huntington's disease, Tourette's syndrome, and dysfunctions of the blood cascade  
 or sexual development disorders, and dysfunctions of the associated gene  
 system including those leading to stroke. The polynucleotides may be used  
 as diagnostic reagents through detecting mutations in the associated gene  
 and for chromosome localization and for tissue expression studies. The  
 polypeptides and polynucleotides may also be used as vaccines.

Sequence 1250 AA;

Query Match 60.7%; Score 37; DB 22; Length 1250;

Best Local Similarity 77.8%; Pred. No. 2.6e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLATGSARQ 9

I:| | | | |

Db 212 cvatnsarq 220

RESULT 12

AAE05251

ID AAE05251 standard; Protein; 1252 AA.

XX AC AAE05251;

XX 12-SEP-2001 (first entry)

XX Mouse Nope (neighbour of punc ell) protein.

XX KW Mouse; Nope; neighbour of punc ell; cytostatic; neuroprotective; vaccine;  
 KW gene therapy; cerebroprotective; colonic cancer; mental retardation;  
 KW tumour suppressor; chromosome 9; transgenic animal; genetic disorder;  
 KW obesity; Bardet-Biedl syndrome; autosomal recessive disorder; retinitis;  
 KW polydactyly; pigmentosa; hypogonadism; hypertension; diabetes mellitus;  
 KW renal anomaly; cardiovascular anomaly.

XX OS Mus musculus.

XX Key Location/Qualifiers

XX Peptide 1..21

XX Protein /label= Signal\_peptide

XX Domain 22..1252

XX Domain /note= "Mature human Nope protein"

XX Domain 22..953

XX Domain /label= Extracellular\_domain

XX Domain 47..127

XX Domain /label= Igl

XX Domain /note= "Immunoglobulin domain 1"

XX Domain 155..218

XX Domain /label= Ig2

FT Domain /note= "Immunoglobulin domain 2"  
 FT 256..318  
 FT /label= Ig3  
 FT /note= "Immunoglobulin domain 3"  
 FT 347..411  
 FT /label= Ig4  
 FT /note= "Immunoglobulin domain 4"  
 FT 429..511  
 FT /label= Fn1  
 FT /note= "Fibronectin-type III domain 1"  
 FT 527..609  
 FT /label= Fn2  
 FT /note= "Fibronectin-type III domain 2"  
 FT 630..725  
 FT /label= Fn3  
 FT /note= "Fibronectin-type III domain 3"  
 FT 750..831  
 FT /label= Fn4  
 FT /note= "Fibronectin-type III domain 4"  
 FT 848..931  
 FT /label= Fn5  
 FT /note= "Fibronectin-type III domain 5"  
 FT 978..1252  
 FT /label= Intracellular\_domain  
 PN WO200149714-A2.  
 XX  
 PD 12-JUL-2001.  
 XX  
 PF 26-OCT-2000; 2000WO-US29698.  
 XX  
 PR 04-JAN-2000; 2000US-0174496.  
 XX  
 PR 19-MAY-2000; 2000US-0205789.  
 XX  
 PA (NEUR-) NEUROSCIENCES RES FOUND INC.  
 XX  
 PI Salbaum JM;  
 XX  
 DR WPI; 2001-441846/47.  
 DR N-PSDB; AAD10021.  
 XX  
 PT Murine Nope polypeptides and nucleic acids useful for preventing,  
 PT diagnosing and treating colonic cancer and Bardet-Biedl syndrome -  
 XX  
 PS Claim 1; Fig 2B; 99pp; English.  
 XX  
 CC The present invention relates to Nope (neighbour of punc ell) which is  
 CC used in the prevention, treatment and diagnosis of diseases associated  
 CC with inappropriate Nope expression such as cancers especially colonic  
 CC cancer and genetic disorders, as Nope is thought to be a tumour  
 CC suppressor. Nope gene is located on chromosome 9 and is used in gene  
 CC therapy. Nope is used as vaccine. Nope gene may be administered to treat  
 CC diseases by rectifying mutations or deletions in a patient's genome that  
 CC affect the activity of Nope by expressing inactive proteins or to  
 CC supplement the patients own production of Nope polypeptides. Nope  
 CC is used to study the expression and function of Nope polypeptides. Nope  
 CC their role in metabolism through the creation of Nope polypeptides and  
 CC models. The anti-Nope antibodies and Nope antagonists may also be used  
 CC to down regulate Nope expression and activity for the treatment of  
 CC Bardet-Biedl syndrome which is an autosomal recessive disorder  
 CC characterised by mental retardation, obesity, polydactyly, retinitis  
 CC pigmentosa and hypogonadism. Patients with Bardet-Biedl syndrome have a  
 CC high incidence of hypertension, diabetes mellitus and renal and  
 CC cardiovascular anomalies. The present sequence is mouse Nope (neighbour  
 CC of punc ell) protein.  
 XX  
 SQ Sequence 1252 AA;

Query Match 60.7%; Score 37; DB 22; Length 1252;  
 Best Local Similarity 77.8%; Pred. NO. 2.6e+02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLATGSARQ 9  
 Db 211 cvatnsarq 219  
 RESULT 13  
 ABG29676  
 ID ABG29676 standard; Protein; 1395 AA.  
 XX  
 AC ABG29676;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #29667.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; RAS93863.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID NO 60035; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABC00010-ABC30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1395 AA;

Query Match 60.7%; Score 37; DB 22; Length 1395;  
 Best Local Similarity 70.0%; Pred. NO. 3e+02;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 ATGSARQLTF 12



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:15:47 ; Search time 21.84 seconds  
(without alignments)  
13.421 Million cell updates/sec

Title: US-08-881-509-8

Perfect score: 61

Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	59.0	428	4	US-09-329-350-33
2	35	57.4	632	4	US-09-232-200-34
3	35	57.4	632	4	US-09-232-200-39
4	35	57.4	632	4	US-09-232-197-34
5	35	57.4	632	4	US-09-232-197-39
6	35	57.4	632	4	US-09-232-201-34
7	35	57.4	632	4	US-09-232-201-39
8	35	57.4	643	4	US-09-232-200-27
9	35	57.4	643	4	US-09-232-200-41
10	35	57.4	643	4	US-09-232-200-53
11	35	57.4	643	4	US-09-232-197-27
12	35	57.4	643	4	US-09-232-197-41
13	35	57.4	643	4	US-09-232-197-53
14	35	57.4	643	4	US-09-232-201-27
15	35	57.4	643	4	US-09-232-201-34
16	35	57.4	643	4	US-09-232-201-53
17	33	54.1	364	4	US-09-232-201-53
18	33	54.1	446	1	US-07-952-800-4
19	33	54.1	456	2	US-08-709-979A-1
20	33	54.1	456	2	US-08-709-979A-11
21	33	54.1	496	4	US-09-147-009-8
22	33	54.1	506	4	US-09-232-191-9
23	33	54.1	506	4	US-09-232-200-9
24	33	54.1	506	4	US-09-232-200-95
25	33	54.1	506	4	US-09-232-197-9
26	33	54.1	506	4	US-09-232-197-95
27	33	54.1	506	4	US-09-232-201-9

28	33	54.1	506	4	US-09-232-201-95
29	33	54.1	525	4	US-08-764-870-7
30	33	54.1	525	4	US-08-980-115-7
31	33	54.1	533	1	US-07-952-800-2
32	33	54.1	632	4	US-09-232-200-35
33	33	54.1	632	4	US-09-232-197-35
34	33	54.1	632	4	US-09-232-201-35
35	33	54.1	643	4	US-09-232-200-45
36	33	54.1	643	4	US-09-232-200-71
37	33	54.1	643	4	US-09-232-200-71
38	33	54.1	643	4	US-09-232-197-45
39	33	54.1	643	4	US-09-232-197-45
40	33	54.1	643	4	US-09-232-197-71
41	33	54.1	643	4	US-09-232-201-42
42	33	54.1	643	4	US-09-232-201-45
43	33	54.1	643	4	US-09-232-201-71
44	33	54.1	2813	3	US-08-896-449A-2
45	33	54.1	2813	3	US-09-132-652-2

## ALIGNMENTS

RESULT 1

US-09-329-350-33

; Sequence 33, Application US/09329350

; Patent No. 6184019

; GENERAL INFORMATION:

; APPLICANT: Miettinen-Oinonen, Arja

; APPLICANT: Londresborough, John

; APPLICANT: Vehmaanper, Jari

; APPLICANT: Haakana, Heli

; APPLICANT: M ntyl, Arja

; APPLICANT: Lantto, Raija

; APPLICANT: Elovainio, Minna

; APPLICANT: Joutsjoki, Vesa

; APPLICANT: Paloheimo, Marja

; APPLICANT: Suominen, Pirkko

; TITLE OF INVENTION: NOVEL CELLULASES, THE GENES ENCODING THEM AND

; TITLE OF INVENTION: USES THEREOF

; NUMBER OF SEQUENCES: 45

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

; STREET: 1100 New York Avenue, N.W., Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/329,350

; FILING DATE: Herewith

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/841,636

; FILING DATE: 30-APR-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/005,335

; FILING DATE: 17-OCT-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/007,926

; FILING DATE: 04-DEC-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/020,840

; FILING DATE: 28-JUN-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/732,181

; FILING DATE: 16-OCT-1996

; PRIOR APPLICATION DATA:

us-08-881-509-8.open.ra1

Thu Jun 6 10:09:38 2002

RESULT 3  
US-09-232-200-39  
; Sequence 39, Application US/09232200A  
; Patent No. 6288213  
; GENERAL INFORMATION:  
; APPLICANT: Stahl, Andreas  
; APPLICANT: Hirsch, David J.  
; APPLICANT: Lodish, Harvey F.  
; APPLICANT: Gimeno, Ruth E.  
; APPLICANT: Tartaglia, Louis A.  
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS  
; FILE REFERENCE: WHI97-21p3MB  
; CURRENT APPLICATION NUMBER: US/09/232,200A  
; CURRENT FILING DATE: 1999-01-14  
; EARLIER APPLICATION NUMBER: 60/071,374  
; EARLIER FILING DATE: 1998-01-15  
; EARLIER APPLICATION NUMBER: 60/093,491  
; EARLIER FILING DATE: 1998-07-20  
; EARLIER APPLICATION NUMBER: 60/110,941  
; EARLIER FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 105  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 39  
; LENGTH: 632  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-232-200-39

Query Match 57.4%; Score 35; DB 4; Length 632;  
Best Local Similarity 58.3%; Pred. No. 60;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CLATGSARQLTF 12  
| | | | |  
Db 157 CLATTSRRLVLF 168

RESULT 4  
US-09-232-197-34  
; Sequence 34, Application US/09232197A  
; Patent No. 6300096  
; GENERAL INFORMATION:  
; APPLICANT: Stahl, Andreas  
; APPLICANT: Hirsch, David J.  
; APPLICANT: Lodish, Harvey F.  
; APPLICANT: Gimeno, Ruth E.  
; APPLICANT: Tartaglia, Louis A.  
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS  
; FILE REFERENCE: WHI97-21p3MA  
; CURRENT APPLICATION NUMBER: US/09/232,197A  
; CURRENT FILING DATE: 1999-01-14  
; EARLIER APPLICATION NUMBER: 60/071,374  
; EARLIER FILING DATE: 1998-01-15  
; EARLIER APPLICATION NUMBER: 60/093,491  
; EARLIER FILING DATE: 1998-07-20  
; EARLIER APPLICATION NUMBER: 60/110,941  
; EARLIER FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 105  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 34  
; LENGTH: 632  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-232-197-34

Query Match 57.4%; Score 35; DB 4; Length 632;  
Best Local Similarity 58.3%; Pred. No. 60;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 CLATGSARQLTF 12  
| | | | |

APPLICATION NUMBER: PCT/FT96/00550  
FILING DATE: 17-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Shea Jr., Timothy  
REGISTRATION NUMBER: 41,306  
REFERENCE/DOCKET NUMBER: 1716.0510006/MAC/TJS  
TELEPHONE: (202)371-2600  
TELEFAX: (202)371-2540  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 428 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Melanocarpus albusmyces  
STRAIN: ALKO4237  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1-428  
OTHER INFORMATION: /label= 50K-cellulase  
US-09-329-350-33

Query Match 59.0%; Score 36; DB 4; Length 428;  
Best Local Similarity 70.0%; Pred. No. 25;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CLATGSARQL 10  
| | | | |  
Db 332 CAATGAARYL 341

RESULT 2  
US-09-232-200-34  
; Sequence 34, Application US/09232200A  
; Patent No. 6288213  
; GENERAL INFORMATION:  
; APPLICANT: Stahl, Andreas  
; APPLICANT: Hirsch, David J.  
; APPLICANT: Lodish, Harvey F.  
; APPLICANT: Gimeno, Ruth E.  
; APPLICANT: Tartaglia, Louis A.  
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS  
; FILE REFERENCE: WHI97-21p3MB  
; CURRENT APPLICATION NUMBER: US/09/232,200A  
; CURRENT FILING DATE: 1999-01-14  
; EARLIER APPLICATION NUMBER: 60/071,374  
; EARLIER FILING DATE: 1998-01-15  
; EARLIER APPLICATION NUMBER: 60/093,491  
; EARLIER FILING DATE: 1998-07-20  
; EARLIER APPLICATION NUMBER: 60/110,941  
; EARLIER FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 105  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 34  
; LENGTH: 632  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-232-200-34

Query Match 57.4%; Score 35; DB 4; Length 632;  
Best Local Similarity 58.3%; Pred. No. 60;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 CLATGSARQLTF 12  
| | | | |  
Db 157 CLATTSRRLVLF 168



Db 157 CLTTSRARALVF 168

RESULT 5  
US-09-232-197-39  
; Sequence 39, Application US/09232197A  
; Patent No. 6300096  
; GENERAL INFORMATION:  
; APPLICANT: Stahl, Andreas  
; APPLICANT: Hirsch, David J.  
; APPLICANT: Lodish, Harvey F.  
; APPLICANT: Gimeno, Ruth E.  
; APPLICANT: Tartaglia, Louis A.  
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS  
; FILE REFERENCE: WHI97-21p3MA  
; CURRENT APPLICATION NUMBER: US/09/232,197A  
; CURRENT FILING DATE: 1999-01-14  
; EARLIER APPLICATION NUMBER: 60/071,374  
; EARLIER FILING DATE: 1998-01-15  
; EARLIER APPLICATION NUMBER: 60/093,491  
; EARLIER FILING DATE: 1998-07-20  
; EARLIER APPLICATION NUMBER: 60/110,941  
; EARLIER FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 105  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 39  
; LENGTH: 632  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-232-197-39

Query Match 57.4%; Score 35; DB 4; Length 632;  
Best Local Similarity 58.3%; Pred. No. 60;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CLATGSARQLTF 12  
||| ||| |

Db 157 CLTTSRARALVF 168

RESULT 6  
US-09-232-201-34  
; Sequence 34, Application US/09232201A  
; Patent No. 6348321  
; GENERAL INFORMATION:  
; APPLICANT: Stahl, Andreas  
; APPLICANT: Hirsch, David J.  
; APPLICANT: Lodish, Harvey F.  
; APPLICANT: Gimeno, Ruth E.  
; APPLICANT: Tartaglia, Louis A.  
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS  
; FILE REFERENCE: WHI97-21p3MC  
; CURRENT APPLICATION NUMBER: US/09/232,201A  
; CURRENT FILING DATE: 1999-01-14  
; EARLIER APPLICATION NUMBER: 60/071,374  
; EARLIER FILING DATE: 1998-01-15  
; EARLIER APPLICATION NUMBER: 60/093,491  
; EARLIER FILING DATE: 1998-07-20  
; EARLIER APPLICATION NUMBER: 60/110,941  
; EARLIER FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 105  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 34  
; LENGTH: 632  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-232-201-34

Query Match 57.4%; Score 35; DB 4; Length 632;  
Best Local Similarity 58.3%; Pred. No. 60;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CLATGSARQLTF 12  
||| ||| |

Db 157 CLTTSRARALVF 168

RESULT 7  
US-09-232-201-39  
; Sequence 39, Application US/09232201A  
; Patent No. 6348321  
; GENERAL INFORMATION:  
; APPLICANT: Stahl, Andreas  
; APPLICANT: Hirsch, David J.  
; APPLICANT: Lodish, Harvey F.  
; APPLICANT: Gimeno, Ruth E.  
; APPLICANT: Tartaglia, Louis A.  
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS  
; FILE REFERENCE: WHI97-21p3MC  
; CURRENT APPLICATION NUMBER: US/09/232,201A  
; CURRENT FILING DATE: 1999-01-14  
; EARLIER APPLICATION NUMBER: 60/071,374  
; EARLIER FILING DATE: 1998-01-15  
; EARLIER APPLICATION NUMBER: 60/093,491  
; EARLIER FILING DATE: 1998-07-20  
; EARLIER APPLICATION NUMBER: 60/110,941  
; EARLIER FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 105  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 39  
; LENGTH: 632  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-232-201-39

Query Match 57.4%; Score 35; DB 4; Length 632;  
Best Local Similarity 58.3%; Pred. No. 60;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CLATGSARQLTF 12  
||| ||| |

Db 157 CLTTSRARALVF 168

RESULT 8  
US-09-232-200-27  
; Sequence 27, Application US/09232200A  
; Patent No. 6288213  
; GENERAL INFORMATION:  
; APPLICANT: Stahl, Andreas  
; APPLICANT: Hirsch, David J.  
; APPLICANT: Lodish, Harvey F.  
; APPLICANT: Gimeno, Ruth E.  
; APPLICANT: Tartaglia, Louis A.  
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS  
; FILE REFERENCE: WHI97-21p3MB  
; CURRENT APPLICATION NUMBER: US/09/232,200A  
; CURRENT FILING DATE: 1999-01-14  
; EARLIER APPLICATION NUMBER: 60/071,374  
; EARLIER FILING DATE: 1998-01-15  
; EARLIER APPLICATION NUMBER: 60/093,491  
; EARLIER FILING DATE: 1998-07-20  
; EARLIER APPLICATION NUMBER: 60/110,941  
; EARLIER FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 105  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 27  
; LENGTH: 643  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-232-200-27

us-08-881-509-8.open.ra1

Thu Jun 6 10:09:38 2002

```

US-09-232-200-53
Query Match      57.4%; Score 35; DB 4; Length 643;
Best Local Similarity 58.3%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CLATGSARQLTF 12
Db 168 CLTTSRARALVF 179

RESULT 11
US-09-232-197-27
; Sequence 27 Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MA
; CURRENT APPLICATION NUMBER: US/09/232,197A
; EARLIER FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-232-197-27

Query Match      57.4%; Score 35; DB 4; Length 643;
Best Local Similarity 58.3%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CLATGSARQLTF 12
Db 168 CLTTSRARALVF 179

RESULT 12
US-09-232-197-41
; Sequence 41 Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MA
; CURRENT APPLICATION NUMBER: US/09/232,197A
; EARLIER FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 41
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-232-197-41

US-09-232-200-41
Query Match      57.4%; Score 35; DB 4; Length 643;
Best Local Similarity 58.3%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CLATGSARQLTF 12
Db 168 CLTTSRARALVF 179

RESULT 9
US-09-232-200-41
; Sequence 41 Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MB
; CURRENT APPLICATION NUMBER: US/09/232,200A
; EARLIER FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-232-200-41

Query Match      57.4%; Score 35; DB 4; Length 643;
Best Local Similarity 58.3%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CLATGSARQLTF 12
Db 168 CLTTSRARALVF 179

RESULT 10
US-09-232-200-53
; Sequence 53 Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MB
; CURRENT APPLICATION NUMBER: US/09/232,200A
; EARLIER FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-232-200-53

```

; LENGTH: 643  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-232-197-41

Query Match 57.4%; Score 35; DB 4; Length 643;  
Best Local Similarity 58.3%; Pred. No. 61;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CLATGSAROLTF 12  
|| | | | |  
Db 168 CLTTSRARALVF 179

## RESULT 13

US-09-232-197-53  
; Sequence 53, Application US/09232197A  
; Patent No. 6300096

; GENERAL INFORMATION:  
; APPLICANT: Stahl, Andreas  
; APPLICANT: Hirsch, David J.  
; APPLICANT: Lodish, Harvey F.  
; APPLICANT: Gimeno, Ruth E.  
; APPLICANT: Tartaglia, Louis A.  
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS  
; FILE REFERENCE: WHI97-21p3MA  
; CURRENT APPLICATION NUMBER: US/09/232,197A  
; EARLIER FILING DATE: 1999-01-14  
; EARLIER APPLICATION NUMBER: 60/071,374  
; EARLIER FILING DATE: 1998-01-15  
; EARLIER APPLICATION NUMBER: 60/093,491  
; EARLIER FILING DATE: 1998-07-20  
; EARLIER APPLICATION NUMBER: 60/110,941  
; EARLIER FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 105  
; SOFTWARE: Fast-SEQ for Windows Version 3.0  
; SEQ ID NO 53  
; LENGTH: 643  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-232-197-53

Query Match 57.4%; Score 35; DB 4; Length 643;  
Best Local Similarity 58.3%; Pred. No. 61;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CLATGSAROLTF 12  
|| | | | |  
Db 168 CLTTSRARALVF 179

## RESULT 14

US-09-232-201-27  
; Sequence 27, Application US/09232201A  
; Patent No. 6348321

; GENERAL INFORMATION:  
; APPLICANT: Stahl, Andreas  
; APPLICANT: Hirsch, David J.  
; APPLICANT: Lodish, Harvey F.  
; APPLICANT: Gimeno, Ruth E.  
; APPLICANT: Tartaglia, Louis A.  
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS  
; FILE REFERENCE: WHI97-21p3MC  
; CURRENT APPLICATION NUMBER: US/09/232,201A  
; EARLIER FILING DATE: 1999-01-14  
; EARLIER APPLICATION NUMBER: 60/071,374  
; EARLIER FILING DATE: 1998-01-15  
; EARLIER APPLICATION NUMBER: 60/093,491  
; EARLIER FILING DATE: 1998-07-20  
; EARLIER APPLICATION NUMBER: 60/110,941  
; EARLIER FILING DATE: 1998-12-04

; NUMBER OF SEQ ID NOS: 105  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 27  
; LENGTH: 643  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-232-201-27

Query Match 57.4%; Score 35; DB 4; Length 643;  
Best Local Similarity 58.3%; Pred. No. 61;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CLATGSAROLTF 12  
|| | | | |  
Db 168 CLTTSRARALVF 179

## RESULT 15

US-09-232-201-41  
; Sequence 41, Application US/09232201A  
; Patent No. 6348321  
; GENERAL INFORMATION:  
; APPLICANT: Stahl, Andreas  
; APPLICANT: Hirsch, David J.  
; APPLICANT: Lodish, Harvey F.  
; APPLICANT: Gimeno, Ruth E.  
; APPLICANT: Tartaglia, Louis A.  
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS  
; FILE REFERENCE: WHI97-21p3MC  
; CURRENT APPLICATION NUMBER: US/09/232,201A  
; EARLIER FILING DATE: 1999-01-14  
; EARLIER APPLICATION NUMBER: 60/071,374  
; EARLIER FILING DATE: 1998-01-15  
; EARLIER APPLICATION NUMBER: 60/093,491  
; EARLIER FILING DATE: 1998-07-20  
; EARLIER APPLICATION NUMBER: 60/110,941  
; EARLIER FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 105  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 41  
; LENGTH: 643  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-232-201-41

Query Match 57.4%; Score 35; DB 4; Length 643;  
Best Local Similarity 58.3%; Pred. No. 61;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CLATGSAROLTF 12  
|| | | | |  
Db 168 CLTTSRARALVF 179

Search completed: June 6, 2002, 06:15:48  
Job time: 32 sec

Thu Jun 6 10:09:38 2002

us-08-881-509-8.open.rai

Page 6



### RESULT

PT0639  
T-cell receptor beta chain V-D-J region (111-1AA) - mouse (fragment)  
T-cell receptor beta chain V-D-J region (111-1AA) - mouse (fragment)  
Species: Mus musculus (house mouse)  
Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
C:Accession: PT0639  
R:Feeney, A.J.  
J. Exp. Med. 174, 115-124, 1991  
A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
A:Reference number: PT0509; MUID:91277601  
A:Accession: PT0639  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-8 <FEE>  
A:Experimental source: newborn thymus, strain BALB/c  
A:Keywords: T-cell receptor

```

Query Match      34.4%; Score 21; DB 2; Length 8;
Best Local Similarity 66.7%; Pred. No. 2.8e+05;
Matches         4: Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

3 ATGSAR 8  
: || ||  
3 STGGAR 8

## RESULT 4

P00783  
NADH dehydrogenase (EC 1.6.99.3) 30K chain - fava bean mitochondrion (fragment)  
Alternate names: complex I 30K chain; NADH-ubiquinone reductase 30K chain  
Species: mitochondrion Vicia faba (fava bean)  
Date: 03-May-1994 #sequence\_revision 07-Oct-1994 #text\_change 17-Mar-1999  
Accession: P00783

C:Accession: S.; Boutry, M.  
R:Letter: S.; 102, 435-443, 1993  
P:Plant Physiol  
A:Title: Purification and preliminary characterization of mitochondrial complex I (NADH: ubiquinone oxidoreductase)  
A:Reference number: PQ0775; PMID:94151437  
A:Accession: PQ0783  
A:Molecule type: protein  
A:Residues: 1-10 <LET>  
C:Comment: Complex I, mitochondrial NADH-ubiquinone reductase, is the first of the three subunits of the mitochondrial electron transfer chain, ranging from 5K to 75k.  
C:Comment: This enzyme catalyzes electron transfer from endogenous NADH to ubiquinone by a series of redox reactions.  
C:Genetics: mitochondrial  
A:Genome: mitochondrial  
A:Transfer: mitochondrial; oxidoreductase.

Query Match	34.4%	Score 21;	DB 2;	Length 10;
Best Local Similarity	44.4%	Pred. NO. 1e+03;		
Matches	4;	Conservative	3;	Mismatches
			2;	Indels
			0;	Gaps
			0;	

QY	3	ATGSARQLT	11
		:	
Dh	1	ATEAAKHIT	9

DEC 11 1955

S65728  
 hemoglobin, extracellular, chain dl - earthworm (*Lumbricus terrestris*) (fragment)  
 C:Species: *Lumbricus terrestris* (common earthworm)  
 C:Date: 06-Dec-1996 #sequence\_revision 13-Mar-1997 #text\_change 13-Mar-1997  
 C:Accession: S65728  
 R:Fushitani, K.; Higashiyama, K.; Asao, M.; Hosokawa, K.  
 Biochim. Biophys. Acta 1292, 273-280, 1996  
 A:Title: Characterization of the constituent polypeptides of the extracellular hemoglobin  
 A:Reference number: S65721; MUID:96176855  
 A:Accession: S65728  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-10 <PUS>

Query Match 34.4%; Score 21; DB 2; Length 10;  
Best Local Similarity 50.0%; Pred. No. 1e+03;  
Matches 4; Conservative 1; Mismatches 3; Indels

QY	1	CLATGSAR	8
			:
Db	2	CLVTESLK	9

RESULT 6

PT0212  
T-cell receptor alpha chain V-J region (4-1-E.2) - mouse (fragment)  
C.Species: Mus musculus (house mouse)  
C.Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 30-1  
C.Accession: PT0212  
R.Nakano, N.; Kikutani, H.; Nishimoto, H.; Kishimoto, T.  
J. Exp. Med. 173, 1091-1097, 1991  
A.Title: T cell receptor V gene usage of islet beta cell-reactive T  
A.Reference number: 91217621  
A.Accession: PT0212  
A.Molecule type: mRNA  
A.Residues: 1-10 <NAK>  
C.Keywords: T-cell receptor

Query Match	34.4%;	Score 21;	DB 2;	Length 10;
Best Local Similarity	40.0%;	Pred. No. 1e+03;		
Matches	4;	Conservative	1;	Mismatches 5;
			Indels	0;
			Gaps	0;

QY 1 CLATGSARQL 10  
| | | : |  
pb 1 CAVAGGADRL 10

7  
RESULTS

PT0568  
T-cell receptor beta chain V-D-J region (141-1C) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-1  
C:Accession: PT0568  
R:Feeney, A.J.  
J. Exp. Med. 174, 115-124, 1991  
A:Title: Functional sequences of fetal T cell receptor beta chains  
A:Reference number: PT0509; MUID:91277601  
A:Accession: PT0568  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-6 <FEE>  
A:Experimental source: day 19 fetal thymus, strain BALB/c  
A:Keywords: T-cell receptor

Query Match	32.8%	Score 20;	DB 2;	Length 6;
Best Local Similarity	66.7%	Pred. No. 2.8e+05;		
Matches	4;	Conservative	1;	Mismatches
			1;	Indels
				Gaps
				0;

QY . 3 ATGSAR 8  
| : | | |  
pb 1 ASGDAR 6

8  
TJLSA

R08996  
 S08996  
 hyperthrealesemic hormone II - oriental cockroach  
 N:Alternate names: Pea-CAH-II  
 C:Species: *Blattella orientalis* (oriental cockroach)  
 C:Date: 30-Jun-1992 #sequence\_revision 24-Oct-1997 #text\_change 31-Oct-1997  
 C:Accession: S08996  
 R:Caede, G.; Rinehart, K.L.  
 Biol. Chem. Hoppe-Seyler 371, 345-354, 1990  
 A:Title: Primary structures of hyperthrealesemic neuropeptides isolated from the cor

entails and of the stick insect Extatosoma tiaratum assigned by tandem fast atom bombardment  
 A:Reference number: S08995; MUID:90253659  
 A:Accession: S08996

A:Molecule type: protein  
 A:Residues: 1-8 <GAE>  
 A:Note: the amino-terminal residue forms pyrrolidone carboxylic acid; therefore, we have  
 C:Superfamily: adipokinetic hormone  
 C:Keywords: amidated carboxyl end; hormone; neuropeptide; pyroglutamic acid  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
 F:8/Modified site: amidated carboxyl end (Trp) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
 Db 1 QLTF 4

RESULT 9

adipokinetic hormone II - American cockroach  
 N:Alternate names: neuropeptide M-II; periplanetin CC-1  
 C:Species: Periplaneta americana (American cockroach)  
 C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 07-May-1999  
 C:Accession: B49823; A05170  
 R:Scarborough, R.M.; Jamieson, G.C.; Kalish, F.; Kramer, S.J.; McEnroe, G.A.; Miller, C.  
 Proc. Natl. Acad. Sci. U.S.A. 81, 5575-5579, 1984  
 A:Title: Isolation and primary structure of two peptides with cardioacceleratory and hypokinetic activities from the American cockroach  
 A:Reference number: B49823; MUID:84298179  
 A:Accession: B49823  
 A:Molecule type: protein  
 A:Residues: 1-8 <SCA>  
 R:Witten, J.L.; Schaffer, M.H.; O'Shea, M.; Cook, J.C.; Hemling, M.E.; Rinehart Jr., K.I.  
 Biochem. Biophys. Res. Commun. 124, 350-358, 1984  
 A:Title: Structures of two cockroach neuropeptides assigned by fast atom bombardment mass spectrometry  
 A:Reference number: A90118; MUID:85046530  
 A:Accession: A05170

A:Molecule type: protein  
 A:Residues: 'E', 2-8 <WIT>  
 C:Superfamily: adipokinetic hormone  
 C:Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic acid  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
 F:8/Modified site: amidated carboxyl end (Trp) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
 Db 1 QLTF 4

RESULT 10

B44960  
 neuropeptide Led-CC-II - Colorado potato beetle  
 C:Species: Leptinotarsa decemlineata (Colorado potato beetle)  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
 C:Accession: B44960  
 R:Gaede, G.; Kellner, R.

Peptides 10; 1287-1289, 1989  
 A:Title: The metabolic neuropeptides of the corpus cardiacum from the potato beetle and their role in the regulation of feeding behavior  
 A:Reference number: A44960; MUID:90160053  
 A:Accession: B44960

A:Molecule type: protein  
 A:Residues: 1-8 <GAE>  
 C:Superfamily: adipokinetic hormone  
 C:Keywords: blocked carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic acid  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

F:8/Modified site: blocked carboxyl end (Trp) (probably amidated) #status experimental  
 Query Match 32.8%; Score 20; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
 Db 1 QLTF 4

RESULT 11

A33995  
 adipokinetic hormone - black horse fly  
 C:Species: Tabanus atratus (black horse fly)  
 C:Date: 23-Mar-1990 #sequence\_revision 23-Mar-1990 #text\_change 31-Oct-1997  
 C:Accession: A33995  
 R:Jaffe, H.; Raina, A.K.; Riley, C.T.; Fraser, B.A.; Nachman, R.J.; Vogel, V.W.; Zhan  
 Proc. Natl. Acad. Sci. U.S.A. 86, 8161-8164, 1989  
 A:Title: Primary structure of two neuropeptide hormones with adipokinetic and hypokinetic activities from the black horse fly  
 A:Reference number: A33995; MUID:90046758  
 A:Accession: A33995

A:Molecule type: protein  
 A:Residues: 1-8 <JAF>  
 C:Superfamily: adipokinetic hormone  
 C:Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic acid  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status predicted  
 F:8/Modified site: amidated carboxyl end (Trp) #status predicted

Query Match 32.8%; Score 20; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
 Db 1 QLTF 4

RESULT 12

S11545  
 adipokinetic hormone - nestling-sucking blowfly  
 C:Species: Protophormia terraenovae (nestling-sucking blowfly)  
 C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 31-Oct-1997  
 C:Accession: S11545  
 R:Gaede, G.; Wilps, H.; Kellner, R.  
 Biochem. J. 269, 309-313, 1990  
 A:Title: Isolation and structure of a novel charged member of the red-pigment-concentrating protein family from the blowfly Protophormia terraenovae  
 A:Reference number: S11545; MUID:90351345  
 A:Accession: S11545

A:Molecule type: protein  
 A:Residues: 1-8 <GAE>  
 C:Superfamily: adipokinetic hormone  
 C:Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic acid  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
 F:8/Modified site: amidated carboxyl end (Trp) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
 Db 1 QLTF 4

RESULT 13

A24244  
 adipokinetic hormone - bollworm  
 N:Alternate names: Hez-AKH

Thu Jun 6 10:09:34 2002

C;Species: Hypertrehalosemic factor II lacks the tryptophan modification.  
 C;Comment: This peptide raises hemolymph levels of trehalose in the cockroach *Periplaneta*  
 C;Superfamily: adipokinetic hormone  
 C;Keywords: amidated carboxyl end; glycoprotein; hormone; neuropeptide; pyroglutamic  
 F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
 F;8/Binding site: carbohydrazide (Trp) (covalent) #status experimental  
 F;10/Modified site: amidated carboxyl end (Thr) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
 ||||  
 Db 1 QLTF 4

Search completed: June 6, 2002, 06:27:51  
 Job time: 119 sec

C;Species: *Heliothis zea* (bollworm, corn earworm, tomato fruitworm)  
 C;Date: 31-Mar-1988 #sequence\_revision 23-Mar-1995 #text\_change 31-Oct-1997  
 C;Accession: A24244  
 R;Jaffe, H.; Raina, A.K.; Riley, C.T.; Fraser, B.A.; Holman, G.M.; Wagner, R.M.; Ridgway  
 Biochem. Biophys. Res. Commun. 135, 622-626, 1986  
 A;Title: Isolation and primary structure of a peptide from the corpora cardiaca of *Heliothis zea*  
 A;Reference number: A24244; MUID:86186794  
 A;Accession: A24244

A;Molecule type: protein  
 A;Residues: 1-9 <DAF>  
 C;Superfamily: adipokinetic hormone  
 C;Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic  
 F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
 F;9/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
 ||||  
 Db 1 QLTF 4

## RESULT 14

A29477  
 diuretic neuropeptide F1 - migratory locust  
 C;Species: *Locusta migratoria* (migratory locust)  
 C;Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 31-Dec-1993  
 C;Accession: A29477  
 R;Proulx, J.P.; Miller, C.A.; Li, J.P.; Carney, R.L.; Girardie, A.; Delaage, M.; Schooley  
 Biochem. Biophys. Res. Commun. 149, 180-186, 1987  
 A;Title: Identification of an arginine vasopressin-like diuretic hormone from *Locusta migratoria*  
 A;Reference number: A29477; MUID:88077077  
 A;Accession: A29477  
 A;Molecule type: protein  
 A;Residues: 1-9 <PKO>  
 A;Note: two neuropeptides, F1 and F2, were identified. F2 is an antiparallel dimer of F1  
 C;Keywords: neuropeptide

Query Match 32.8%; Score 20; DB 2; Length 9;  
 Best Local Similarity 50.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CLATGSAR 8  
 ||||  
 Db 1 CLITNCPR 8

## RESULT 15

JC1416  
 hypertrehalosemic hormone I - stick insect (*Carausius morosus*)  
 N;Alternate names: neuropeptide Cam-HrH-I  
 N;Contains: hypertrehalosemic factor II  
 C;Species: *Carausius morosus*  
 C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
 C;Accession: JC1416; S07157  
 R;Gaede, G.; Kellner, R.; Rinehart, K.L.; Proefke, M.L.  
 Biochem. Biophys. Res. Commun. 189, 1303-1309, 1992  
 A;Title: A tryptophan-substituted member of the AKH/RPCH family isolated from a stick insect  
 A;Reference number: JC1416; MUID:93129188  
 A;Accession: JC1416  
 A;Molecule type: protein  
 A;Residues: 1-10 <GAEI>  
 R;Gaede, G.; Rinehart Jr., K.L.  
 Biol. Chem. Hoppe-Seyler 368, 67-75, 1987  
 A;Title: Primary structure of the hypertrehalosemic factor II from the corpus cardiacum  
 A;Reference number: S07157; MUID:87157103  
 A;Accession: S07157  
 A;Molecule type: protein  
 A;Residues: 'Z', 2-10 <GAE2>



entails and of the stick insect Extatosoma tiaratum assigned by tandem fast atom bombard

A:Reference number: S08995; MUID:90253659  
 A:Accession: S08996  
 A:Molecule type: protein  
 A:Residues: 1-8 <GAE>  
 A:Note: the amino-terminal residue forms pyrrolidone carboxylic acid; therefore, we have  
 C:Superfamily: adipokinetic hormone  
 C:Keywords: amidated carboxyl end; hormone; neuropeptide; pyroglutamic acid  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
 F:8/Modified site: amidated carboxyl end (Trp) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
 ||||  
 Db 1 QLTF 4

RESULT 9

B49823  
 adipokinetic hormone II - American cockroach  
 N:Alternate names: neuropeptide M-II; periplanetin CC-1  
 C:Species: Periplaneta americana (American cockroach)  
 C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 07-May-1999  
 C:Accession: B49823; A05170  
 R:Scarborough, R.M.; Jamieson, G.C.; Kalish, F.; Kramer, S.J.; McEnroe, G.A.; Miller, C.  
 Proc. Natl. Acad. Sci. U.S.A. 81, 5575-5579, 1984  
 A:Title: Isolation and primary structure of two peptides with cardioacceleratory and hyp

A:Reference number: A49823; MUID:84298179  
 A:Accession: B49823  
 A:Molecule type: protein  
 A:Residues: 1-8 <SCA>  
 R:Witten, J.L.; Schaffer, M.H.; O'Shea, M.; Cook, J.C.; Hemling, M.E.; Rinehart Jr., K.L.  
 Biochem. Biophys. Res. Commun. 124, 350-358, 1984  
 A:Title: Structures of two cockroach neuropeptides assigned by fast atom bombardment mas

A:Reference number: A90118; MUID:85046530  
 A:Accession: A05170  
 A:Molecule type: protein  
 A:Residues: 'E', 2-8 <WTP>  
 C:Superfamily: adipokinetic hormone  
 C:Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
 F:8/Modified site: amidated carboxyl end (Trp) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
 ||||  
 Db 1 QLTF 4

RESULT 10

B44960  
 neuropeptide Led-CC-II - Colorado potato beetle  
 C:Species: Leptinotarsa decemlineata (Colorado potato beetle)  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
 C:Accession: B44960  
 R:Gaede, G.; Kellner, R.  
 Peptides 10, 1287-1289, 1989  
 A:Title: The metabolic neuropeptides of the corpus cardiacum from the potato beetle and

A:Reference number: A44960; MUID:90160053  
 A:Accession: B44960  
 A:Molecule type: protein  
 A:Residues: 1-8 <GAE>  
 C:Superfamily: adipokinetic hormone  
 C:Keywords: blocked carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

F:8/Modified site: blocked carboxyl end (Trp) (probably amidated) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
 ||||  
 Db 1 QLTF 4

RESULT 11

A33995  
 adipokinetic hormone - black horse fly  
 C:Species: Tabanus atratus (black horse fly)  
 C:Date: 23-Mar-1990 #sequence\_revision 23-Mar-1990 #text\_change 31-Oct-1997  
 C:Accession: A33995  
 R:Jaffe, H.; Raina, A.K.; Riley, C.T.; Fraser, B.A.; Nachman, R.J.; Vogel, V.W.; Zhan  
 Proc. Natl. Acad. Sci. U.S.A. 86, 8161-8164, 1989  
 A:Title: Primary structure of two neuropeptide hormones with adipokinetic and hypotri

A:Reference number: A33995; MUID:90046758  
 A:Accession: A33995  
 A:Molecule type: protein  
 A:Residues: 1-8 <JAF>  
 C:Superfamily: adipokinetic hormone  
 C:Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyrogluta  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status predicted  
 F:8/Modified site: amidated carboxyl end (Trp) #status predicted

Query Match 32.8%; Score 20; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
 ||||  
 Db 1 QLTF 4

RESULT 12

S11545  
 adipokinetic hormone - nestling-sucking blowfly  
 C:Species: Protophormia terraenovae (nestling-sucking blowfly)  
 C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 31-Oct-1997  
 C:Accession: S11545  
 R:Gaede, G.; Wilps, H.; Kellner, R.  
 Biochem. J. 269, 309-313, 1990  
 A:Title: Isolation and structure of a novel charged member of the red-pigment-concent

erraeovae (Diptera).  
 A:Reference number: S11545; MUID:90351345  
 A:Accession: S11545  
 A:Molecule type: protein  
 A:Residues: 1-8 <GAE>  
 C:Superfamily: adipokinetic hormone  
 C:Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyrogluta  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
 F:8/Modified site: amidated carboxyl end (Trp) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
 ||||  
 Db 1 QLTF 4

RESULT 13

A24244  
 adipokinetic hormone - bollworm  
 N:Alternate names: Hez-AKH

us-08-881-509-8.closed.rpr

Thu Jun 6 10:09:34 2002

C;Comment: Hypertrehalosemic factor II lacks the tryptophan modification.  
 C;Date: 31-Mar-1988 #sequence\_revision 23-Mar-1995 #text\_change 31-Oct-1997  
 C;Accession: A24244  
 R;Jaife, H.; Raina, A.K.; Riley, C.T.; Fraser, B.A.; Holman, G.M.; Wagner, R.M.; Ridgway  
 Biochem. Biophys. Res. Commun. 135, 622-628, 1986  
 A;Title: Isolation and primary structure of a peptide from the corpora cardiaca of Helic  
 A;Reference number: A24244; MUID:86186794  
 A;Accession: A24244  
 A;Molecule type: protein  
 A;Residues: 1-9 <YAF>  
 C;Superfamily: adipokinetic hormone  
 C;Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
 F:9/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 32.8%; Score 20; DB 2; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
 ||||  
 Db 1 QLTF 4

Search completed: June 6, 2002, 06:27:51  
 Job time: 119 sec

Query Match 32.8%; Score 20; DB 2; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
 ||||  
 Db 1 QLTF 4

## RESULT 14

A29477  
 diuretic neuropeptide F1 - migratory locust  
 C;Species: Locusta migratoria (migratory locust)  
 C;Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 31-Dec-1993  
 C;Accession: A29477  
 R;Proux, J.P.; Miller, C.A.; Li, J.P.; Carney, R.L.; Girardie, A.; Delaage, M.; Schooley  
 Biochem. Biophys. Res. Commun. 149, 180-186, 1987  
 A;Title: Identification of an arginine vasopressin-like diuretic hormone from Locusta m  
 A;Reference number: A29477; MUID:88077077  
 A;Accession: A29477  
 A;Molecule type: protein  
 A;Residues: 1-9 <PRO>  
 A;Note: two neuropeptides, F1 and F2, were identified. F2 is an antiparallel dimer of F1  
 C;Keywords: neuropeptide

Query Match 32.8%; Score 20; DB 2; Length 9;  
 Best Local Similarity 50.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CLATGSAR 8  
 ||||  
 Db 1 CLITNCPR 8

## RESULT 15

JC1416  
 hypertrehalosemic hormone I - stick insect (Carausius morosus)  
 N;Alternate names: neuropeptide Cam-HrTH-I  
 N;Contains: hypertrehalosemic factor II  
 C;Species: Carausius morosus  
 C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
 C;Accession: JC1416; S07157  
 R;Gaede, G.; Kellner, R.; Rinehart, K.L.; Proefke, M.L.  
 Biochem. Biophys. Res. Commun. 189, 1303-1309, 1992  
 A;Title: A tryptophan-substituted member of the AKH/RPCH family isolated from a stick in  
 A;Reference number: JC1416; MUID:93129188  
 A;Accession: JC1416  
 A;Molecule type: protein  
 A;Residues: 1-10 <GAEI>  
 R;Gaede, G.; Rinehart Jr., K.L.  
 Biol. Chem. Hoppe-Seyler 368, 67-75, 1987  
 A;Title: Primary structure of the hypertrehalosemic factor II from the corpus cardiacum  
 A;Reference number: S07157; MUID:87157103  
 A;Accession: S07157  
 A;Molecule type: protein  
 A;Residues: 1-10 <GAEI>

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:26:37 ; Search time 10.17 Seconds  
(without alignments)  
45.687 Million cell updates/sec

Title: US-08-881-509-8  
Perfect score: 61  
Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 463

Minimum DB seq length: 0  
Maximum DB seq length: 12

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	32.8	8	1	AKH_TABAT
2	20	32.8	8	1	HTF1_PERAM
3	20	32.8	9	1	DNF1_LOCHI
4	20	32.8	10	1	HTF2_CARMO
5	20	32.8	10	1	HTF1_HELZE
6	20	32.8	10	1	HTF1_TABAT
7	17	27.9	10	1	AKHX_LOCHI
8	16	26.2	12	1	CALM_TETH
9	15	24.6	7	1	UH11_RAT
10	15	24.6	8	1	HTF1_TENMO
11	15	24.6	8	1	RPCH_PANBO
12	15	24.6	8	1	UPAA_HUMAN
13	15	24.6	9	1	AL11_CARMA
14	15	24.6	9	1	LITR_PHYRO
15	15	24.6	10	1	GON2_CHEPR
16	15	24.6	10	1	VEG6_BACSU
17	15	24.6	11	1	CA41_LITCI
18	15	24.6	11	1	CA42_LITCI
19	15	24.6	11	1	RANC_RANPI
20	14	23.0	8	1	ALL4_CYDPO
21	14	23.0	9	1	HUTU_KLEAP
22	14	23.0	9	1	XYLA_STRSQ
23	14	23.0	10	1	GATU_HUMAN
24	14	23.0	10	1	URE3_MORMO
25	14	23.0	11	1	MORN_HUMAN
26	13	21.3	8	1	LPM5_STAEP
27	13	21.3	9	1	FAR9_ASCSU
28	13	21.3	9	1	LMT3_LOCHI
29	13	21.3	10	1	MOSQ_CLXTA
30	13	21.3	10	1	TKUL_UREUN
31	13	21.3	12	1	TIN2_HOPTI
32	12	19.7	8	1	AKHG_GRYBI
33	12	19.7	8	1	AKH_LIBAU

34	12	19.7	8	1	AKH_MELML	P25423 melolontha
35	12	19.7	8	1	HTF1_PERAM	P04548 periplaneta
36	12	19.7	9	1	CONO_CONST	P05487 conus stria
37	12	19.7	9	1	DSIP_RABIT	P01158 cryptolaqus
38	12	19.7	10	1	FARC_CALVO	P41867 calliphora
39	12	19.7	10	1	HTF1_ROMMI	P18110 romalea mic
40	12	19.7	10	1	HTF1_ROMMI	P10939 nauphoeta c
41	12	19.7	10	1	PSBF_CAPAN	OQ3367 capsicua an
42	12	19.7	10	1	RCA_PINPS	P01084 pinus pinas
43	12	19.7	11	1	CA21_LITCI	P82087 litoria cit
44	12	19.7	11	1	CA22_LITCI	P82088 litoria cit
45	12	19.7	11	1	TIN4_HOPTI	P82654 hoplobatrach

## ALIGNMENTS

RESULT 1  
AKH\_TABAT  
ID AKH\_TABAT STANDARD; PRT; 8 AA.  
AC P14595;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DE Adipokinetic hormone (AKH) (Dipteran corpora cardiaca factor I)  
DE (DCC I).  
OS Tabanus atratus (Horse fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Tabanomorpha;  
OC Tabanidae; Tabanus.  
OX NCBI\_TaxID=7207;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Corpora cardiaca;  
RX MEDLINE=90046758; Pubmed=2813385;  
RA Jaffe H., Raina A.K., Riley C.T., Fraser B.A., Nachman R.J.,  
Vogel V.W., Zhang Y.-S., Hayes D.K.;  
RT "Primary structure of two neuro peptide hormones with adipokinetic and  
hypotrehalosemic activity isolated from the corpora cardiaca of horse  
flies (Diptera).";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:8161-8164(1989).  
CC -!- FUNCTION: THIS HORMONE, RELEASED FROM CELLS IN THE CORPORA  
CARDIACA AFTER THE BEGINNING OF FLIGHT, CAUSES RELEASE OF  
DIGLYCERIDES FROM THE FAT BODY AND THEN STIMULATES THE FLIGHT  
MUSCLES TO USE THESE DIGLYCERIDES AS AN ENERGY SOURCE.  
CC -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.  
CC PIR; A33995; A33995.  
DR InterPro: IPR002047; AKH.  
DR PROSITE: PS00256; AKH; 1.  
KW Neuropeptide; Amidation; Flight.  
FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
FT MOD\_RES 8 8 AMIDATION.  
SQ SEQUENCE 8 AA; 949 MW; 86786771A9D1A736 CRC64;

Query Match 32.8%; Score 20; DB 1; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
|  
|  
|  
|  
Db 1 QLTF 4

RESULT 2  
HTF2\_PERAM  
ID HTF2\_PERAM STANDARD; PRT; 8 AA.  
AC P04549;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DE Hypertrehalosemic factor II (Neuropeptide M-II) (Periplanetin CC-2)  
(Pea-CAH-II) (Leb-CC-II) (Hypertrehalosemic neuropeptide II).

OS Periplaneta americana (American cockroach),  
 OS Leptinotarsa decemlineata (Colorado potato beetle), and  
 OS Blatta orientalis (Oriental cockroach).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;  
 OC Blattodea; Blattidae; Periplaneta.  
 OX NCBI\_TaxID=6978, 7539, 6976;  
 RN [1]  
 RN SEQUENCE.  
 RP SPECIES=P.americana;  
 RC MEDLINE=85046530; PubMed=6548628;  
 RA Witten J.L., Schaffer M.H., O'Shea M., Cook J.C., Hemling M.E.,  
 RA Rinehart K.L. Jr.;  
 RT "Structures of two cockroach neuropeptides assigned by fast atom  
 RT bombardment mass spectrometry.";  
 RL Biochem. Biophys. Res. Commun. 124:350-358(1984).  
 RN [2]  
 RN SEQUENCE.  
 RP SPECIES=P.americana;  
 RC MEDLINE=84298179; PubMed=6591205;  
 RX Scarbrough R.M., Jamieson G.C., Kalish F., Kramer S.J., McEnroe G.A.,  
 RA Miller C.A., Schooley D.A.;  
 RT "Isolation and primary structure of two peptides with  
 RT cardioacceleratory and hyperglycemic activity from the corpora  
 RT cardiaca of Periplaneta americana.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 81:5575-5579(1984).  
 RN [3]  
 RN SEQUENCE.  
 RP SPECIES=L.declineata; TISSUE=Corpora cardiaca;  
 RC MEDLINE=90160053; PubMed=2576128;  
 RX Gaede G., Kellner R.;  
 RT "The metabolic neuropeptides of the corpus cardiaca from the potato  
 RT beetle and the American cockroach are identical.";  
 RL Peptides 10:1287-1289(1989).  
 RN [4]  
 RN SEQUENCE.  
 RP SPECIES=B.orientalis; TISSUE=Corpora cardiaca;  
 RC MEDLINE=90253659; PubMed=2340112;  
 RX Gaede G., Rinehart K.L. Jr.;  
 RT "Primary structures of hypertrehalosaemic neuropeptides isolated from  
 RT the corpora cardiaca of the cockroaches Leucophaea maderae,  
 RT Gromphadorhina portentosa, Blattella germanica and Blatta orientalis  
 RT and of the stick insect Extatosoma tiaratum assigned by tandem fast  
 RT atom bombardment mass spectrometry.";  
 RL Biol. Chem. Hoppe-Seyler 371:345-354(1990).  
 CC -1- FUNCTION: HYPERTREHALOSAEMIC FACTORS ARE NEUROPEPTIDES THAT  
 CC ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS  
 CC THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).  
 CC -1- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.  
 DR PIR: A05170; A05170.  
 DR PIR: S08996; S08996.  
 DR PIR: B44960; B44960.  
 DR PIR: B49823; B49823.  
 DR InterPro; IPR002047; AKH.  
 DR PROSITE; PS00256; AKH; 1.  
 KW Neuropeptide; Amidation.  
 FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 FT MOD\_RES 8 8 AMIDATION.  
 FT SEQUENCE 8 AA; 1006 MW; 86745771A9D1A736 CRC64;  
 SQ  
 Query Match 32.8%; Score 20; DB 1; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 9 QLTFF 12  
 DB 1 QLTFF 4  
 RESULT 3  
 DNFL\_LOCMI STANDARD; PRT; 9 AA.  
 ID DNFL\_LOCMI

AC P16339;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Locupressin (Diuretic neuropeptide F1/F2).  
 OS Locusta migratoria (Migratory locust).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;  
 OC Acridomorpha; Acridoidea; Acrididae; Locusta.  
 OX NCBI\_TaxID=7004;  
 RN [1]  
 RN SEQUENCE.  
 RP TISSUE=Suboesophageal ganglion; and Thoracic ganglion;  
 RC MEDLINE=88077077; PubMed=3689410;  
 RX Proux J.P., Miller C.A., Li J.P., Carney R.L., Girardie A.,  
 RA Delaage M., Schooley D.A.;  
 RT "Identification of an arginine vasopressin-like diuretic hormone from  
 RT Locusta migratoria.";  
 RL Biochem. Biophys. Res. Commun. 149:180-186(1987).  
 CC -1- FUNCTION: DIURETIC HORMONE.  
 CC -1- SUBUNIT: F2 IS AN ANTIPARALLEL DISULFIDE LINKED DIMER OF F1.  
 CC -1- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.  
 DR PIR: A29477; A29477.  
 DR InterPro; IPR000981; Neurohypophys\_horm.  
 DR Pfam; PF00220; hormone4; 1.  
 DR PROSITE; PS00264; NEUROHYPOPHYS\_HORM; 1.  
 KW Hormone; Neuropeptide; Amidation.  
 FT DISULFID 1 6 IN F1.  
 FT DISULFID 1 1 INTERCHAIN (WITH C-6') (IN F2).  
 FT DISULFID 6 6 INTERCHAIN (WITH C-1') (IN F2).  
 FT MOD\_RES 9 9 AMIDATION.  
 FT SEQUENCE 9 AA; 976 MW; 56EB176EB451A057 CRC64;  
 SQ  
 Query Match 32.8%; Score 20; DB 1; Length 9;  
 Best Local Similarity 50.0%; Pred. No. 1e+05;  
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 CLATGSAR 8  
 DB 1 CLITNCPR 8  
 RESULT 4  
 HTF2\_CARMO STANDARD; PRT; 10 AA.  
 ID HTF2\_CARMO  
 AC P11385; (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Hypertrehalosaemic factor II (HTF-II) (HPTH-II) (Hypertrehalosaemic  
 DE neuropeptide II).  
 OS Carausius morosus (Indian stick insect), and  
 OS Extatosoma tiaratum (Stick insect).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Orthopteroidea; Phasmatodea; Heteronemiidae;  
 OC Carausius.  
 OX NCBI\_TaxID=7022, 7024;  
 RN [1]  
 RN SEQUENCE.  
 RP SPECIES=C.morosus; TISSUE=Corpora cardiaca;  
 RC MEDLINE=87157103; PubMed=3828078;  
 RX Gaede G., Rinehart K.L. Jr.;  
 RT "Primary structure of the hypertrehalosaemic factor II from the  
 RT corpus cardiaca of the Indian stick insect, Carausius morosus,  
 RT determined by fast atom bombardment mass spectrometry.";  
 RL Biol. Chem. Hoppe-Seyler 368:67-75(1987).  
 RN [2]  
 RN SEQUENCE.  
 RP SPECIES=E.tiaratum; TISSUE=Corpora cardiaca;  
 RC MEDLINE=90253659; PubMed=2340112;  
 RX Gaede G., Rinehart K.L. Jr.;  
 RT "Primary structures of hypertrehalosaemic neuropeptides isolated from

RT the corpora cardiaca of the cockroaches *Leucophaea maderae*,  
 RT *Gromphadorhina portentosa*, *Blattella germanica* and *Blatta orientalis*  
 RT and of the stick insect *Extatosoma tiaratum* assigned by tandem fast  
 RT atom bombardment mass spectrometry.";  
 RL Biol. Chem. Hoppe-Seyler 371:345-354 (1990).  
 RN [3]

RP CARBOHYDRATE-LINKAGE SITE.  
 RC SPECIES-C. MOROSUS; TISSUE-Corpora cardiaca;  
 RX MEDLINE-93129188; PubMed-1482345;  
 RA Gaede G., Kellner R., Rinehart K.L. Jr., Proefke M.L.;  
 RT "A tryptophan-substituted member of the AKH/RPCH family isolated from  
 RL a stick insect corpus cardiacum.";  
 RL Biochem. Biophys. Res. Commun. 189:1303-1309 (1992).  
 CC ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS  
 CC THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).  
 CC -1- MASS SPECTROMETRY: MW=1308.61; METHOD-FAB.  
 CC -1- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.  
 DR PIR; S07157; S07157.  
 DR PIR; S09138; S09138.  
 DR InterPro: IPR002047; AKH.  
 DR PROSITE; PS00256; AKH; 1.  
 KW Neuropeptide; Amidation; Glycoprotein.  
 FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 FT CARBOHYD 8 8 C-LINKED (MAN) (PROBABLE).  
 FT MOD\_RES 10 10 AMIDATION.  
 FT MOD\_RES 10 10  
 SQ SEQUENCE 10 AA; 1164 MW; 9B9036745771A9D1 CRC64;

Query Match 32.8%; Score 20; DB 1; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
 DB 1 QLTF 4

RESULT 5  
 HTF\_HELZE  
 ID HTF\_HELZE STANDARD; PRT; 10 AA.  
 AC P16353;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Hypertrehalosemic hormone (Hex-HRTH).  
 OS Heliothis zea (corn earworm) (Bollworm).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Noctuidae; Noctuidae; Heliothinae; Helicoverpa.  
 OX NCBI\_TaxID=7113;  
 RN [1]  
 RP SEQUENCE.

RC TISSUE-Corpora cardiaca;  
 RX MEDLINE-88326324; PubMed-3415690;  
 RA Jaffe H., Raina A.K., Riley C.T., Fraser B.A., Bird T.G.,  
 RA Tseng C.M., Zhang Y.S., Hayes D.K.;  
 RT "Isolation and primary structure of a neuropeptide hormone from  
 RL Heliothis zea with hypertrehalosemic and adipokinetic activities.";  
 RL Biochem. Biophys. Res. Commun. 155:344-350 (1988).  
 CC ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS  
 CC THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).  
 CC -1- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.  
 DR PIR; A31571; A31571.  
 DR InterPro: IPR002047; AKH.  
 DR PROSITE; PS00256; AKH; 1.  
 KW Neuropeptide; Amidation.  
 FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 FT MOD\_RES 10 10 AMIDATION.  
 FT MOD\_RES 10 10  
 SQ SEQUENCE 10 AA; 1096 MW; 8E70367865A5B9D1 CRC64;

Query Match 32.8%; Score 20; DB 1; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
 DB 1 QLTF 4

RESULT 6  
 HTF\_TABAT  
 ID HTF\_TABAT STANDARD; PRT; 10 AA.  
 AC P14596;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Hypertrehalosemic factor (HOTH) (Dipteran corpora cardiaca factor II)  
 DE (DCC II).  
 OS Tabanus atratus (Horse fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Tabanomorpha;  
 OC Tabanidae; Tabanus.  
 OX NCBI\_TaxID=7207;  
 RN [1]  
 RP SEQUENCE.

RC TISSUE-Corpora cardiaca;  
 RX MEDLINE-90046758; PubMed-2813385;  
 RA Jaffe H., Raina A.K., Riley C.T., Fraser B.A., Nachman R.J.,  
 RA Vogel V.W., Zhang Y.-S., Hayes D.K.;  
 RT "Primary structure of two neuropeptide hormones with adipokinetic and  
 RT hypotrehalosemic activity isolated from the corpora cardiaca of horse  
 RT flies (Diptera)."  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:8161-8164 (1989).  
 CC ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS  
 CC THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).  
 CC -1- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.  
 DR PIR; B33995; B33995.  
 DR InterPro: IPR002047; AKH.  
 DR PROSITE; PS00256; AKH; 1.  
 KW Neuropeptide; Amidation.  
 FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 FT MOD\_RES 10 10 AMIDATION.  
 FT MOD\_RES 10 10  
 SQ SEQUENCE 10 AA; 1169 MW; 916036786771A9D1 CRC64;

Query Match 32.8%; Score 20; DB 1; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12  
 DB 1 QLTF 4

RESULT 7  
 AKHX\_LOCMI  
 ID AKHX\_LOCMI STANDARD; PRT; 10 AA.  
 AC P81626;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Peptide hormone.  
 OS Locusta migratoria (Migratory locust).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;  
 OC Acridomorpha; Acridoidea; Acrididae; Locusta.  
 OX NCBI\_TaxID=7004;  
 RN [1]  
 RP SEQUENCE.

RC TISSUE-Corpora cardiaca;  
 RA Siebert K.J.;  
 RL Submitted (DEC-1998) to the SWISS-PROT data bank.

Thu Jun 6 10:09:35 2002

```

CC -!- FUNCTION: PROBABLY INVOLVED IN THE REGULATION OF LOCUST
CC INTERMEDIARY METABOLISM, BEHAVIOR AND/OR DEVELOPMENT.
CC -!- SIMILARITY: SOME SIMILARITY TO THE AKH / HRTH / RPCH FAMILY.
DR InterPro: IPR002047; AKH.
KW Neuropeptide; Amidation.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 10 10 AMIDATION.
FT MOD_RES 10 10 1222 MW; 81BFF67AB415B9D1 CRC64;
SQ SEQUENCE 10 AA; 1222 MW; 81BFF67AB415B9D1 CRC64;

Query Match 27.9%; Score 17; DB 1; Length 10;
Best Local Similarity 75.0%; Pred. No. 1.8e+03;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLTF 12
DB 1 QVTF 4

RESULT 8
CALM_TETH STANDARD; PRT; 12 AA.
AC Q05055;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Calmodulin (Fragment).
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OC NCBI_TaxID=5911;
[1]
SEQUENCE FROM N.A.
MEDLINE=93281388; PubMed=8506136;
Kato M., Hirano M., Takemasa T., Kimura M., Watanabe Y.;
"A micronucleus-specific sequence exists in the 5'-upstream region of
the calmodulin gene in Tetrahymena thermophila.";
RL Nucleic Acids Res. 21:2409-2414(1993).
CC -!- FUNCTION: CALMODULIN MEDIATES THE CONTROL OF A LARGE NUMBER OF
CC ENZYMES BY CA(++) . AMONG THE ENZYMES TO BE STIMULATED BY THE
CC CALMODULIN-CA(++) COMPLEX ARE A NUMBER OF PROTEIN KINASES AND
CC PHOSPHATASES.
CC -!- MISCELLANEOUS: THIS PROTEIN HAS FOUR FUNCTIONAL CALCIUM-BINDING
CC SITES.
CC -!- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: D12774; BAA02239.1; -.
DR HSP: P02593; 2CLN.
DR InterPro: IPR002048; EF-hand.
DR PROSITE: PS00018; EF_HAND; PARTIAL.
KW Calcium-binding; Repeat; Acetylation.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT NON_TER 12 12
FT SEQUENCE 12 AA; 1393 MW; 83F31CD443DB1B01 CRC64;

Query Match 26.2%; Score 16; DB 1; Length 12;
Best Local Similarity 80.0%; Pred. No. 3.3e+03;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 AROLT 11
DB 1 AQLT 5

us-08-881-509-8.closed.rsp

```

```

RESULT 9
UH11_RAT STANDARD; PRT; 7 AA.
AC P56576;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Unknown protein from 2D-page of heart tissue (Spot P11) (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
[1]
SEQUENCE.
RC STRAIN-WISTAR; TISSUE=Heart;
RA Li X.-P., Pleissner K.-P., Scheler C., Regitz-Zagrosek V., Salikov J.,
RA Jungblut P.R.;
RL Submitted (SEP-1998) to the SWISS-PROT data bank.
CC -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC PROTEIN IS: 8.5, ITS MW IS: 42 kDa.
CC UNSURE 2 2 OR A.
FT NON_TER 7 7
FT SEQUENCE 7 AA; 775 MW; 6866DB040DC5A6B0 CRC64;

Query Match 24.6%; Score 15; DB 1; Length 7;
Best Local Similarity 75.0%; Pred. No. 1e+05;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 SARQ 9
DB 2 SARE 5

RESULT 10
HTF_TENMO STANDARD; PRT; 8 AA.
AC P25419;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Hypertrehalosemic factor (HOTH) (Hypertrehalosemic neuropeptide).
OS Tenebrio molitor (Yellow mealworm), and
OS Zophobas rugipes.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OC NCBI_TaxID=7067, 7075;
[1]
SEQUENCE.
RC TISSUE=Corpora cardiaca;
RX MEDLINE=90341081; PubMed=2381871;
RA Gaede G., Rosinski G.;
RT "The primary structure of the hypertrehalosemic neuropeptide from
tenebrionid beetles: a novel member of the AKH/RPCH family.";
RL Peptides 11:455-459(1990).
CC -!- FUNCTION: HYPERTREHALOSEMIC FACTORS ARE NEUROPEPTIDES THAT
CC ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH OF INSECTS).
CC THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).
CC -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR PIR: A43976; A43976.
DR PIR: B43976; B43976.
DR InterPro: IPR002047; AKH.
DR PROSITE: PS00256; AKH; 1.
KW Neuropeptide; Amidation.
FT MOD_RES 1 1
FT MOD_RES 8 8
FT SEQUENCE 8 AA; 1005 MW; 86745775B9C44736 CRC64;

Query Match 24.6%; Score 15; DB 1; Length 8;
Best Local Similarity 75.0%; Pred. No. 1e+05;

```

Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 QLTF 12  
 DB 1 QLN 4

RESULT 11  
 RPCH\_PANBO  
 ID RPCH\_PANBO STANDARD; PRT; 8 AA.  
 AC P08939;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Red pigment concentrating hormone (RPCH).  
 OS Pandanus borealis (Northern red shrimp).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Caridea; Pandaloidea;  
 OC Pandalidae; Pandalus.  
 OX NCBI\_TaxID=6703;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=75054965; PubMed=4433569;  
 RA Ferlund P.;  
 RT "Structure of the red-pigment-concentrating hormone of the shrimp,  
 Pandanus borealis.";  
 RL Biochim. Biophys. Acta 371:304-311(1974).  
 CC -1- FUNCTION: THIS HORMONE ADAPTS THE ANIMAL TO LIGHT BACKGROUNDS BY  
 CC STIMULATING CONCENTRATION OF THE PIGMENT OF ITS RED BODY-  
 CC CHROMATOPHORES.  
 CC -1- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.  
 DR PIR: S07139; S07139.  
 DR InterPro: IPR002047; AKH.  
 DR PROSITE: PS00256; AKH; 1.  
 KW Pigment; Hormone; Amidation.  
 FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 FT MOD\_RES 8 8 AMIDATION.  
 SQ SEQUENCE 8 AA; 948 MW; 86786775B9C44736 CRC64;

Query Match 24.6%; Score 15; DB 1; Length 8;  
 Best Local Similarity 75.0%; Pred. No. 1e+05;  
 Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 QLTF 12  
 DB 1 QLN 4

RESULT 12  
 UPAA\_HUMAN  
 ID UPAA\_HUMAN STANDARD; PRT; 8 AA.  
 AC P30096;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Unknown protein from 2D-page of plasma (Spot 36) (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE.  
 RX TISSUE=Plasma;  
 RX MEDLINE=93092937; PubMed=1459097;  
 RA Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquali C.,  
 RA Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B.,  
 RA Hochstrasser D.F.;  
 RT "Plasma protein map: an update by microsequencing.";  
 RL Electrophoresis 13:707-714(1992).  
 CC -1- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN  
 CC PROTEIN IS: 7. ITS MW IS: 12 KDa.  
 DR SWISS-2DPAGE: P30096; HUMAN.

FT NON\_TER 1 1  
 FT VARIANT 5 5 F -> P.  
 FT FT /FTId=VAR\_000004.  
 SQ SEQUENCE 8 AA; 909 MW; 86677B59D1A72042 CRC64;

Query Match 24.6%; Score 15; DB 1; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LTF 12  
 DB 3 LTF 5

RESULT 13  
 AL11\_CARMA  
 ID AL11\_CARMA STANDARD; PRT; 9 AA.  
 AC P81814;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Carcinus maenas.  
 OS Carcinus maenas (Common shore crab) (Green crab).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
 OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
 OX NCBI\_TaxID=6759;  
 RN [1]  
 RP SEQUENCE.  
 RX TISSUE=Cerebral ganglion, and Thoracic ganglion;  
 RX MEDLINE=98121193; PubMed=9461295;  
 RA Duve H., Johnsen A.H., Maestri J.-L., Scott A.G., Jaros P.P.,  
 RA Thorpe A.;  
 RT "Isolation and identification of multiple neuropeptides of the  
 RT allatostatin superfamily in the shore crab Carcinus maenas.";  
 RL Eur. J. Biochem. 250:727-734(1997).  
 CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
 KW Neuropeptide; Amidation; Multigene family.  
 FT MOD\_RES 9 9 AMIDATION.  
 SQ SEQUENCE 9 AA; 927 MW; 832D79CDB46D861 CRC64;

Query Match 24.6%; Score 15; DB 1; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ATG 5  
 DB 1 ATG 3

RESULT 14  
 LITR\_PHYRO  
 ID LITR\_PHYRO STANDARD; PRT; 9 AA.  
 AC P08946;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Rhodel-litorin.  
 OS Phyllomedusa rohdei (Rohde's leaf frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
 OC Phyllomedusa.  
 OX NCBI\_TaxID=8394;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=85127560; PubMed=3838283;  
 RA Barria D., Erspamer G.F., Simmaco M., Bossa F., Melchiorri P.,  
 RA Erspamer V.;  
 RT "Rhodel-litorin: a new peptide from the skin of Phyllomedusa rohdei.";

Job time: 197 sec

RL FEBS Lett. 182:53-56(1985).  
CC -|- SUBCELLULAR LOCATION: Secreted.  
CC -|- SIMILARITY: BELONGS TO THE BOMBESIN/NEUROMEDIN B/RANATENSIN  
CC FAMILY.  
DR PIR; S07241; S07241.  
DR InterPro; IPR000874; Bombesin.  
DR Pfam; PF02044; Bombesin; 1.  
DR PROSITE; PS00257; BOMBESIN; 1.  
KW Bombesin family; Amidation.  
FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
FT MOD\_RES 9 9 AMIDATION.  
SQ SEQUENCE 9 AA; 1090 MW; 4ECCCE1861ADC377 CRC64;

Query Match 24.6%; Score 15; DB 1; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 3 ATG 5  
DB 4 ATG 6

RESULT 15  
ID GON2\_CHEPR STANDARD; PRT: 10 AA.  
AC P80678;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Gonadoliberin II (Gonadotropin-releasing hormone II) (GNRH-II)  
DE (Luliberin II).  
OS Chelyosoma productum.  
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Phlebobranchia;  
OC Corellidae; Chelyosoma.  
OX NCBI\_TaxID=71177;  
RN [1]  
RP MEDLINE=96413669; PubMed=8816823;  
RA Powell J.F.F., Reska-Skinner S.M., Prakash M.O., Fischer W.H.,  
RA Park M., Rivier J.E., Craig A.G., Mackie G.O., Sherwood N.M.;  
RT "Two new forms of gonadotropin-releasing hormone in a protochordate  
RT and the evolutionary implications."  
RL Proc. Natl. Acad. Sci. U.S.A. 93:10461-10464(1996).  
CC -|- FUNCTION: STIMULATES THE SECRETION OF BOTH LUTEINIZING AND  
CC FOLLICLE-STIMULATING HORMONES.  
CC -|- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.  
CC -|- SUBCELLULAR LOCATION: Secreted.  
CC -|- TISSUE SPECIFICITY: GNRH NEURONS LIE WITHIN BLOOD SINUSES CLOSE TO  
CC THE GONODUCTS AND GONADS IN BOTH JUVENILES AND ADULTS, IMPLYING  
CC THAT THE NEUROPEPTIDE IS RELEASED INTO THE BLOODSTREAM.  
CC -|- MASS SPECTROMETRY: MW=1117.52; METHOD=MALDI.  
CC -|- SIMILARITY: BELONGS TO THE GNRH FAMILY.  
DR InterPro; IPR002012; Gnrh.  
DR PROSITE; PS00473; GNRH; 1.  
KW Hormone; Amidation.  
FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
FT DISULFID 6 6 INTERCHAIN.  
FT MOD\_RES 10 10 AMIDATION (BY SIMILARITY).  
FT MOD\_RES 10 10 284B38D1EEB735A3 CRC64;  
SQ SEQUENCE 10 AA; 1135 MW;

Query Match 24.6%; Score 15; DB 1; Length 10;  
Best Local Similarity 60.0%; Pred. No. 4.3e+03;  
Matches 3; Conservative 0; Mismatches 2; Indels 0; Caps 0;

QY 1 CLATG 5  
DB 6 CHAPG 10



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:27:32 ; Search time 23.79 seconds  
(without alignments)  
87.261 Million cell updates/sec

Title: US-08-881-509-8  
Perfect score: 61  
Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 1842

Minimum DB seq length: 0  
Maximum DB seq length: 12

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL19.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23	37.7	9	11 Q9QZAB	Q9QZAB mus musculus
2	21	34.4	8	11 P70243	P70243 mus musculus
3	21	34.4	12	10 Q945C3	Q945C3 cryptosporidi
4	19	31.1	8	11 Q60615	Q60615 mus musculus
5	19	31.1	11	11 Q9R0K9	Q9R0K9 mus musculus
6	18	29.5	8	2 Q9Q258	Q9Q258 synchococc
7	18	29.5	10	11 Q9R0Z0	Q9R0Z0 mus musculus
8	18	29.5	12	10 Q02319	Q02319 pinus sylve
9	18	29.5	12	10 Q02320	Q02320 pinus sylve
10	18	29.5	12	10 Q38715	Q38715 arachis hyp
11	17	27.9	7	2 Q47505	Q47505 escherichia
12	17	27.9	11	4 Q9UNL8	Q9UNL8 homo sapien
13	17	27.9	12	12 Q10421	Q10421 influenza a
14	16	26.2	9	7 Q31415	Q31415 gallus gall
15	16	26.2	11	5 P82700	P82700 leucophaea
16	16	26.2	12	4 Q9UBJ5	Q9UBJ5 homo sapien

17	16	26.2	12	10 Q9SYT4	Q9SYT4 arabidopsis
18	16	26.2	12	10 Q41856	Q41856 zea mays (m
19	15	24.6	8	6 P82929	P82929 bos taurus
20	15	24.6	8	12 Q89965	Q89965 polyomaviru
21	15	24.6	9	12 Q9PYK1	Q9PYK1 simian viru
22	15	24.6	9	12 Q9IBM8	Q9IBM8 simian viru
23	15	24.6	10	10 P81899	P81899 prunus dulc
24	15	24.6	11	3 Q9UR95	Q9UR95 pichia angu
25	15	24.6	11	11 Q99JC3	Q99JC3 rattus sp.
26	15	24.6	11	12 Q40974	Q40974 cauliflower
27	15	24.6	11	13 Q90735	Q90735 gallus gall
28	15	24.6	12	2 P83054	P83054 bacteroides
29	15	24.6	12	3 Q9UR28	Q9UR28 filobasidio
30	15	24.6	12	3 Q9UR22	Q9UR22 cryptococcu
31	15	24.6	12	6 Q9TRU9	Q9TRU9 bos taurus
32	15	24.6	12	9 Q9TLJ4	Q9TLJ4 bacterioph
33	15	24.6	12	10 P82329	P82329 pisum sativ
34	15	24.6	12	12 Q69232	Q69232 bovine herp
35	14	23.0	7	13 Q42564	Q42564 fugu rubrip
36	14	23.0	8	2 Q44463	Q44463 rhizobiacea
37	14	23.0	9	6 Q9XT05	Q9XT05 macropus ru
38	14	23.0	10	5 P82222	P82222 bombyx mori
39	14	23.0	11	5 P82698	P82698 leucophaea
40	14	23.0	11	6 Q9TRX0	Q9TRX0 sus scrofa
41	14	23.0	11	7 Q77513	Q77513 oreochromis
42	14	23.0	11	11 Q9QXK6	Q9QXK6 mus musculu
43	14	23.0	11	11 Q60807	Q60807 mus musculu
44	14	23.0	12	2 Q9R7F1	Q9R7F1 staphylococ
45	14	23.0	12	6 Q9XT42	Q9XT42 canis famil

## ALIGNMENTS

RESULT 1  
Q9QZAB ID Q9QZAB PRELIMINARY; PRT; 9 AA.  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE C-TYPE LECTIN DCL1 (FRAGMENT).  
GN DCL1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gershi K., Huang X., Tseng S.-Y., Rattis F., Pardoll D., Tsuchiya H.;  
RT "Dendritic cell regulation of DCL1 mRNA expression."  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF192526; AAF04843.1;  
DR MGD; MGI:2136650; Dcl1.  
KW Lectin.  
FT NON\_TER  
SQ SEQUENCE 9 AA; 984 MW; 342161AB172EBAB7 CRC64;  
Query Match 37.7%; Score 23; DB 11; Length 9;  
Best Local Similarity 80.0%; Pred. No. 5.6e+05;  
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CLATG 5  
Db 4 CLETG 8  
RESULT 2  
P70243 ID P70243 PRELIMINARY; PRT; 8 AA.  
AC P70243;  
DT 01-FEB-1997 (TrEMBLrel. 02, Created)

Thu Jun 6 10:09:36 2002

```

DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE SKELETAL MUSCLE-SPECIFIC CALCIUM CHANNEL (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Ophoff R.A.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DE EMBL; X98325; CAA66969.1; -.
FT NON_TER 1
FT NON_TER 8
FT NON_TER 8
SQ SEQUENCE 8 AA; 865 MW; D9C37DDB1861ADDE CRC64;

Query Match 34.4%; Score 21; DB 11; Length 8;
Best Local Similarity 71.4%; Pred. No. 5.6e+05;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATGSARQ 9
Db 2 ATGEAWQ 8

RESULT 3
Q945C3 PRELIMINARY; PRT; 12 AA.
AC Q945C3; 2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE P80 PROTEIN (FRAGMENT).
OS Cryptocodinium cohnii (Dinoflagellate).
OC Eukaryota; Alveolata; Dinophyceae; Gonyaulacales; Cryptecodiniaceae;
OC Cryptecodinium.
OX NCBI_TaxID=2866;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99433238; PubMed=10505419;
RA Ausell J., Soyer-Gobillard M.O., Geraud M.L., Bhaud Y., Baines I.,
RA Preston T., Moreau H.;
RT "Characterization of p80, a novel nuclear and cytoplasmic protein in
RT dinoflagellates."
RL Proctist 150:197-211(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Guillebaud D., Derelle E., Lozano J.C., Bingham S., Moreau H.;
RT "A single TBP-like protein is present in the marine unicellular
RT organism: the dinoflagellate Cryptecodinium cohnii."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF417568; AAL15906.1; -.
FT NON_TER 12
FT NON_TER 12
SQ SEQUENCE 12 AA; 1364 MW; 615BF873FE204414 CRC64;

Query Match 34.4%; Score 21; DB 10; Length 12;
Best Local Similarity 57.1%; Pred. No. 2e+03;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 6 SARQLTF 12
Db 3 SARRILF 9

RESULT 4
Q60615 PRELIMINARY; PRT; 8 AA.
ID Q60615
AC Q60615;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE APOLIPROTEIN A-II (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Ko M.S., Wang X., Horton J.H., Hagen M.D., Takahashi N., Maezaki Y.,
RA Nadeau J.H.;
RT "Genetic mapping of 40 cDNA clones on the mouse genome by PCR."
RL Mamm. Genome 5:349-355(1994).
DR EMBL; U05691; AAB60462.1; -.
DR MGI; MGI:88050; Apo2.
DR Lipoprotein.
FT NON_TER 1
FT NON_TER 8
SQ SEQUENCE 8 AA; 720 MW; 783DDDC5B861AB18 CRC64;

Query Match 31.1%; Score 19; DB 11; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.6e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TGSA 7
Db 3 TGSA 6

RESULT 5
Q9R0K9 PRELIMINARY; PRT; 11 AA.
ID Q9R0K9
AC Q9R0K9;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE BCL-2 ASSOCIATED X PROTEIN (FRAGMENT).
GN BAX OR BAX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20035745; PubMed=10570968;
RA Igata E., Inoue T., Ohtani-Fujita N., Sowa Y., Tsujimoto Y., Sakai T.;
RT "Molecular cloning and functional analysis of the murine bax gene
RT promoter."
RL Gene 238:407-415(1999).
DR EMBL; AB029557; BAA82406.1; -.
DR MGI; MGI:99702; Bax.
FT NON_TER 11
FT NON_TER 11
SQ SEQUENCE 11 AA; 1037 MW; 90AAB34E36DB1865 CRC64;

Query Match 31.1%; Score 19; DB 11; Length 11;
Best Local Similarity 66.7%; Pred. No. 4.5e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 GSARQL 10
Db 3 GSQEQL 8

RESULT 6
Q09258 PRELIMINARY; PRT; 8 AA.
ID Q09258
AC Q09258;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

```

```

DE NIFH (FRAGMENT).
GN NIFH.
ON Synechococcus sp. (strain PCC 8801 / RF-1) (Cyanothecae PCC 8801).
OC Bacteria; Cyanobacteria; Chroococcales; Cyanothecae.
OX NCBI_TaxID=41431;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RF-1;
RX MEDLINE=99231861; PubMed=10217509;
RA Huang T.C., Lin R.F., Chu M.K., Chen H.M.;
RT "Organization and expression of nitrogen-fixation genes in the aerobic
RT nitrogen-fixing unicellular cyanobacterium Synechococcus sp. strain
RT RF-1.";
RL Microbiology 145:743-753(1999).
DR EMBL; AF001780; AAC33369.1; -.
FT NON_TER 8
SQ SEQUENCE 8 AA; 985 MW; F16B59CDD046C406 CRC64;

Query Match 29.5%; Score 18; DB 2; Length 8;
Best Local Similarity 60.0%; Pred. No. 5.6e+05;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 8 ROLTF 12
Db 2 RQIAF 6

RESULT 7
Q9R0Z0 Q9R0Z0 PRELIMINARY; PRT; 10 AA.
AC Q9R0Z0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE Cw17 PROTEIN (FRAGMENT).
GN Cw17
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=LIVER;
RX MEDLINE=99287587; PubMed=10360842;
RA Wrehke C., Wiedenmeyer W.R., Schmitt-Wrede H.P., Mincheva A.,
RA Lichke P., Wunderlich F.;
RT "Genomic organization of mouse gene zfp162 (mzf).";
RL DNA Cell Biol. 18:419-428(1999).
DR EMBL; Y14702; CAB45189.1; -.
FT NON_TER 10
SQ SEQUENCE 10 AA; 946 MW; 836D48ADD44DD861 CRC64;

Query Match 29.5%; Score 18; DB 11; Length 10;
Best Local Similarity 60.0%; Pred. No. 6.3e+03;
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LATGS 6
Db 1 MATGA 5

RESULT 8
Q02319 Q02319 PRELIMINARY; PRT; 12 AA.
AC Q02319;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE CHALCONE SYNTHASE (FRAGMENT).
OS Pinus sylvestris (Scots pine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

OC OC
OX NCBI_TaxID=3349;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SEEDLING;
RX MEDLINE=93015948; PubMed=1400374;
RA Schroder G., Schroder J.;
RT "A single change of histidine to glutamine alters the substrate
RT preference of a stilbene synthase.";
RL J. Biol. Chem. 267:20558-20560(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=SEEDLING;
RX MEDLINE=92163014; PubMed=1536925;
RA Fliegmann J., Schroder G., Schanz S., Britsch L., Schroder J.;
RT "Molecular analysis of chalcone and dihydropinosylvin synthase from
RT Scots pine (Pinus sylvestris), and differential regulation of these
RT and related enzyme activities in stressed plants.";
RL Plant Mol. Biol. 18:489-503(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=SEEDLING;
RA Schroder J.;
RL Submitted (AUG-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; L00657; AAA50522.1; -.
FT NON_TER 12
SQ SEQUENCE 12 AA; 1303 MW; DF8E5A1BE0CEB866 CRC64;

Query Match 29.5%; Score 18; DB 10; Length 12;
Best Local Similarity 50.0%; Pred. No. 7.6e+03;
Matches 3; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CLATGS 6
Db 7 CFAGGT 12

RESULT 9
Q02320 Q02320 PRELIMINARY; PRT; 12 AA.
AC Q02320;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE CHALCONE SYNTHASE (FRAGMENT).
OS Pinus sylvestris (Scots pine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=3349;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93015948; PubMed=1400374;
RA Schroder G., Schroder J.;
RT "A single change of histidine to glutamine alters the substrate
RT preference of a stilbene synthase.";
RL J. Biol. Chem. 267:20558-20560(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92163014; PubMed=1536925;
RA Fliegmann J., Schroder G., Schanz S., Britsch L., Schroder J.;
RT "Molecular analysis of chalcone and dihydropinosylvin synthase from
RT Scots pine (Pinus sylvestris), and differential regulation of these
RT and related enzyme activities in stressed plants.";
RL Plant Mol. Biol. 18:489-503(1992).
RN [3]
RP SEQUENCE FROM N.A.
RA Schroder J.;
RL Submitted (AUG-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; L00658; AAA50523.1; -.
FT NON_TER 12
SQ SEQUENCE 12 AA; 1303 MW; EDB35A1BE0CEB871 CRC64;

```

```

[1]
RN  SEQUENCE FROM N.A.
RP  MEDLINE=9609297; PubMed=8522520;
RX  Gonzalez-Pastor J.E., San Millan J.L., Castilla M.A., Moreno F.;
RA  "Structure and organization of plasmid genes required to produce the
RT  translation inhibitor microcin C7.";
RL  J. Bacteriol. 177:7131-7140(1995).
DR  EMBL; X57583; CAA40808.1; -.
KW  Plasmid.
SQ  SEQUENCE 7 AA; 763 MW; 644DD44861B406F0 CRC64;

Query Match 27.9%; Score 17; DB 2; Length 7;
Best Local Similarity 50.0%; Pred. No. 5.6e+05;
Matches 3; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY  2 LATGSA 7
Db  1 MRTGNA 6

RESULT 12
Q9UNL8 PRELIMINARY; PRT; 11 AA.
ID  Q9UNL8
AC  Q9UNL8;
DT  01-MAY-2000 (TREMBlrel. 13, Created)
DT  01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT  01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE  APC2 PROTEIN (FRAGMENT).
GN  APC2.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX  NCBI_TaxID=9606;
[1]
RN  SEQUENCE FROM N.A.
RP  Carr I.M., Markham A.F., Colleta P.L., Wai L., Askham J., Morrison E.,
RA  Meredith D.M.;
RT  "APC2 partial gene sequence.";
RL  Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AF110338; AAD29275.1; -.
FT  NON_TER 1
FT  SEQUENCE 11 AA; 1326 MW; 75881D7BB441EAB4 CRC64;

Query Match 27.9%; Score 17; DB 4; Length 11;
Best Local Similarity 75.0%; Pred. No. 1.1e+04;
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  1 CLAT 4
Db  8 CLQT 11

RESULT 13
O10421 PRELIMINARY; PRT; 12 AA.
ID  O10421
AC  O10421;
DT  01-JUL-1997 (TREMBlrel. 04, Created)
DT  01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT  01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE  NEURAMINIDASE (FRAGMENT).
OS  Influenza A virus (A/South Carolina/1/18 (H1N1)).
OC  Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC  Influenza virus A and B group; Influenza A viruses; Influenza A virus.
OX  NCBI_TaxID=59375;
[1]
RN  SEQUENCE FROM N.A.
RP  STRAIN=A/SOUTH CAROLINA/1/18 (H1N1);
RX  MEDLINE=97218304; PubMed=9065404;
RA  Taubenberger J.K., Reid A.H., Krafft A.E., Bijwaard K.E.,
RA  Fanning T.G.;

```

```

Query Match 29.5%; Score 18; DB 10; Length 12;
Best Local Similarity 50.0%; Pred. No. 7.6e+03;
Matches 3; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY  1 CLATGS 6
Db  7 CFAGGT 12

RESULT 10
Q38715 PRELIMINARY; PRT; 12 AA.
ID  Q38715
AC  Q38715;
DT  01-NOV-1996 (TREMBlrel. 01, Created)
DT  01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT  01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE  RESVERATROL SYNTHASE (FRAGMENT).
OS  Arachis hypogaea (peanut).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC  eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC  Arachis.
OX  NCBI_TaxID=3818;
[1]
RN  SEQUENCE FROM N.A.
RP  Lanz T., Schroder G., Schroder J.;
RA  "Differential regulation of genes for resveratrol synthase in cell
RT  cultures of Arachis hypogaea L.";
RL  Planta 181:169-175(1990).
[2]
RN  SEQUENCE FROM N.A.
RP  MEDLINE=91236783; PubMed=2033084;
RA  Schroder J.;
RT  "The role of cysteines in polyketide synthases. Site-directed
RT  mutagenesis of resveratrol and chalcone synthases, two key enzymes in
RT  different plant specific pathways.";
RL  J. Biol. Chem. 266:9971-9976(1991).
[3]
RN  SEQUENCE FROM N.A.
RP  MEDLINE=93015948; PubMed=1400374;
RA  Schroeder G., Schroeder J.;
RT  "A single change of histidine to glutamine alters the substrate
RT  preference of a stilbene synthase.";
RL  J. Biol. Chem. 267:20558-20560(1992).
RL  EMBL; L00953; AAC41685.1; -.
FT  NON_TER 12
FT  SEQUENCE 12 AA; 1294 MW; C0B35A1BE0CEB866 CRC64;

Query Match 29.5%; Score 18; DB 10; Length 12;
Best Local Similarity 50.0%; Pred. No. 7.6e+03;
Matches 3; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY  1 CLATGS 6
Db  7 CFAGGT 12

RESULT 11
Q47505 PRELIMINARY; PRT; 7 AA.
ID  Q47505
AC  Q47505;
DT  01-NOV-1996 (TREMBlrel. 01, Created)
DT  01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT  01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE  MCCA PROTEIN.
OS  Escherichia coli.
OC  Bacterium; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC  Escherichia.
OX  NCBI_TaxID=562;

```

```

RT "Initial genetic characterization of the 1918 'Spanish' influenza
RL virus.";
RL Science 275:1793-1796(1997).
DR EMBL; U94893; AAC57065.1; -.
FT NON_TER 1
FT NON_TER 12
FT NON_TER 12
SQ SEQUENCE 12 AA; 1248 MW; D13CE0DB8F5862D2 CRC64;

Query Match 27.9%; Score 17; DB 12; Length 12;
Best Local Similarity 40.0%; Pred. No. 1.2e+04;
Matches 2; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CLATG 5
Db 3 CMVVG 7

RESULT 14
Q31415 PRELIMINARY; PRT; 9 AA.
AC Q31415;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-JAN-1999 (TREMBlrel. 09, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE MHC CLASS I ANTIGEN (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Kanki T., Kuwasa N., Sekiya Y., Ichikawa Y.;
RT "Responsive expression of a MHC class I epitope and genes following
RT Marek's disease virus infection.";
RL Submitted (JAN-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; D90399; BAA14395.1; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 9 AA; 859 MW; 8A55A76455B861B5 CRC64;

Query Match 26.2%; Score 16; DB 7; Length 9;
Best Local Similarity 75.0%; Pred. No. 5.6e+05;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 ATGS 6
Db 2 STGS 5

RESULT 15
P82700 PRELIMINARY; PRT; 11 AA.
AC P82700;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PERIVISCERININ-3 (LEM-PVK-3).
OS Leucophaea maderae (Madelra cockroach),
OS Nauphoeta cinerea (Cinereous cockroach) (Gray cockroach),
OS Blaberus craniifer,
OS Blaptica dubia (Argentinian wood cockroach), and
OS Gromphadorina portentosa (Cockroach).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC Blaberoides; Blaberidae; Leucophaea.
OX NCBI_TaxID=6988, 6990, 6982, 132935, 36953;
RN [1]
RP SEQUENCE, FUNCTION, AND MASS SPECTROSCOPY.
RC TISSUE-ABDOMINAL PERISYMPATHETIC ORGANS.
RX MEDLINE=20307624; PubMed=10849006;
RA Predel R., Kellner R., Baggerman G., Steinmetzer T., Schoofs L.;

```

```

RT "Identification of novel periviscerokinins from single neurohaemal
RT release sites in insects. MS/MS fragmentation complemented by Edman
RT degradation.";
RL Eur. J. Biochem. 267:3869-3873(2000).
CC -1- FUNCTION: MEDIATES VISCERAL MUSCLE CONTRACTILE ACTIVITY
CC (MYOTROPIC ACTIVITY).
CC -1- MASS SPECTROMETRY: MW=1146.6; METHOD=MALDI.
KW Neuropeptide; Amidation.
FT MOD_RES 11 11
SQ SEQUENCE 11 AA; 1147 MW; 2F4D9FF2D7605698 CRC64;

```

```

Query Match 26.2%; Score 16; DB 5; Length 11;
Best Local Similarity 37.5%; Pred. No. 1.7e+04;
Matches 3; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 GSARQLTF 12
Db 1 GSSGMIPF 8

```

Search completed: June 6, 2002, 06:30:44  
Job time: 192 sec

us-08-881-509-8.closed.rspt

Thu Jun 6 10:09:36 2002

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:23:47 ; Search time 29.23 Seconds  
(without alignments)  
45.600 Million cell updates/sec

Title: US-08-881-509-8

Perfect score: 61  
Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 158732

Minimum DB seq length: 0  
Maximum DB seq length: 12

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	61	100.0	12	19 AAW47591	T-cell receptor CD
2	28	45.9	9	15 AAR55332	Inhibitor of tyros
3	28	45.9	10	22 AAG95029	Human complementar
4	28	45.9	10	22 AAG95031	Human complementar
5	28	45.9	10	22 AAG95037	Human complementar
6	28	45.9	10	22 AAG95041	Human complementar
7	28	45.9	10	22 AAG95045	Human complementar
8	28	45.9	10	22 AAG95091	Human complementar
9	28	45.9	10	22 AAG95093	Human complementar
10	28	45.9	10	22 AAG95095	Human complementar
11	28	45.9	10	22 AAG95097	Human complementar

12	28	45.9	10	22 AAG95099	Human complementar
13	28	45.9	10	22 AAG95101	Human complementar
14	28	45.9	10	22 AAG95115	Human complementar
15	28	45.9	10	22 AAG95141	Human complementar
16	28	45.9	10	22 AAG95143	Human complementar
17	28	45.9	10	22 AAG95187	Human complementar
18	28	45.9	10	22 AAG95223	Human complementar
19	28	45.9	10	22 AAG95233	Human complementar
20	28	45.9	10	22 AAG95235	Human complementar
21	28	45.9	10	22 AAG95237	Human complementar
22	28	45.9	10	22 AAG95239	Human complementar
23	28	45.9	10	22 AAG95243	Human complementar
24	28	45.9	10	22 AAG95251	Human complementar
25	28	45.9	10	22 AAG95257	Human complementar
26	28	45.9	10	22 AAG95295	Human complementar
27	28	45.9	10	22 AAG95311	Human complementar
28	28	45.9	10	22 AAG95327	Human complementar
29	28	45.9	10	22 AAG95329	Human complementar
30	28	45.9	10	22 AAG95335	Human complementar
31	28	45.9	10	22 AAG95339	Human complementar
32	28	45.9	10	22 AAG95341	Human complementar
33	28	45.9	10	22 AAG95343	Human complementar
34	28	45.9	10	22 AAG95349	Human complementar
35	28	45.9	10	22 AAG95353	Human complementar
36	28	45.9	10	22 AAG95359	Human complementar
37	28	45.9	10	22 AAG95365	Human complementar
38	28	45.9	10	22 AAG95375	Human complementar
39	28	45.9	10	22 AAG95379	Human complementar
40	28	45.9	10	22 AAG95389	Human complementar
41	28	45.9	10	22 AAG95397	Human complementar
42	28	45.9	11	19 AAW64705	Seq ID 43 from WO9
43	27	44.3	10	22 AAG95027	Human complementar
44	27	44.3	10	22 AAG95087	Human complementar
45	27	44.3	10	22 AAG95089	Human complementar

## ALIGNMENTS

RESULT 1  
AAW47591  
ID AAW47591 standard; peptide; 12 AA.  
XX  
AC AAW47591;  
XX  
DT 26-JUN-1998 (first entry)  
XX  
DE T-cell receptor CDR3 alpha-region.  
XX  
KW Alpha-region; human; T-cell receptor; TCR; diagnosis; monitoring;  
KW prevention; therapy; tumour disease; renal cell carcinoma;  
KW CDR3.  
XX  
OS Homo sapiens.  
XX  
PN DE19625191-A1.  
XX  
PD 02-JAN-1998.  
XX  
PF 24-JUN-1996; 96DE-1025191.  
XX  
PR 24-JUN-1996; 96DE-1025191.  
XX  
PA (BOEF ) BOEHRINGER MANNHEIM GMBH.  
XX  
PI Schendel D;  
XX  
DR WPI; 1998-053442/06.  
XX  
PT N-PSDB; AAV18708.  
XX  
PT Human T-cell receptor nucleic acids and poly:peptide(s) - for  
diagnosis or therapy, especially of renal cell carcinoma

Thu Jun 6 10:09:31 2002

PS Example 1; Page 17; 30pp; German.

XX The present sequence is the CDR3 alpha-region of a human  
 CC T-cell receptor (TCR), which can be used in the diagnosis,  
 CC monitoring, prevention and therapy of a tumour disease,  
 CC specifically renal cell carcinoma.

XX Sequence 12 AA;

Query Match 100.0%; Score 61; DB 19; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 4.8e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLATGSARQLTF 12  
 |||||  
 Db 1 clatgsarqltf 12

RESULT 2

AAAR55332  
 ID AAR55332 standard; peptide; 9 AA.

XX AC AAR55332;

XX DT 30-JAN-1995 (first entry)

XX DE Inhibitor of tyrosine kinase.

XX PTX; signal transduction; proliferation; cancer; psoriasis; viral;  
 KW inflammatory; allergic; cardiovascular; diseases.

XX OS Synthetic.

XX PN WO9411392-A.

XX PD 26-MAY-1994.

XX PF 03-NOV-1993; 93WO-US10610.

XX PR 06-NOV-1992; 92US-0973136.

XX PR 27-OCT-1993; 93US-0139913.

XX PA (WARN ) WARNER LAMBERT CO.

XX PI Dobrusin EM, Fry DW, Mcnamara DJ, Singh J;

XX DR WPI; 1994-183419/22.

XX PT New peptide inhibitors of protein tyrosine kinase - contain D-Tyr  
 or tetrafluoro-Tyr residue, for treating proliferative diseases,  
 PT viral infection, inflammation etc.

XX PS Claim 7; Page 36; 44pp; English.

XX CC The sequence is that of a protein tyrosine kinase inhibitor contg. D-  
 Tyr or tetrafluoro-Tyr residues. The peptide can be used to inhibit  
 CC TR-mediated signal transduction and is useful for controlling  
 CC proliferative diseases, e.g. cancer, psoriasis and to treat viral,  
 CC inflammatory, allergic and cardiovascular diseases.  
 CC See also AAR55302-39.

XX SQ Sequence 9 AA;

Query Match 45.9%; Score 28; DB 15; Length 9;  
 Best Local Similarity 75.0%; Pred. No. 6.4e-05;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LATGSARQ 9  
 |||||  
 Db 1 laegsare 8

RESULT 3

AAAG95029  
 ID AAG95029 standard; Peptide; 10 AA.

XX AC AAG95029;

XX DT 18-SEP-2001 (first entry)

XX DE Human complementary peptide, SEQ ID NO: 1223.

XX KW Human; complementary peptide; ligand; drug discovery; drug design.

XX OS Homo sapiens.

XX PN WO200142277-A2.

XX PD 14-JUN-2001.

XX PF 13-DEC-2000; 2000WO-GB04776.

XX PR 13-DEC-1999; 99GB-0029464.

XX PA (PROT-) PROTEOM LTD.

XX PI Roberts GW, Heal JR;

XX DR WPI; 2001-408419/43.

XX A set of peptide ligands consisting of specific complementary peptides  
 PT to proteins encoded by genes of the human genome, useful in an assay  
 PT for screening and identifying of one or more novel peptides which are  
 PT drug candidates or pro-drugs -

XX Example 4; Page 215; 646pp; English.

XX The invention relates to a set of complementary peptide ligands  
 CC generated from the human genome. The complementary peptides  
 CC interact with their relevant target proteins encoded in the human  
 CC genome. They can be used as reagents in drug discovery and as lead  
 CC ligands to facilitate drug design and development. The present  
 CC sequence is a complementary peptide provided in the specification.

XX SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;  
 Best Local Similarity 55.6%; Pred. No. 79;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLATGSARQ 9

Db 1 cartgsgr 9

RESULT 4

AAAG95031  
 ID AAG95031 standard; Peptide; 10 AA.

XX AC AAG95031;

XX DT 18-SEP-2001 (first entry)

XX DE Human complementary peptide, SEQ ID NO: 1225.

XX KW Human; complementary peptide; ligand; drug discovery; drug design.  
 XX OS Homo sapiens.

XX PN WO200142277-A2.

XX PD 14-JUN-2001.

XX



PF 13-DEC-2000; 2000WO-GB04776.  
XX  
PR 13-DEC-1999; 99GB-0029464.  
XX  
PA (PROT-) PROTEOM LTD.  
XX  
PI Roberts GW, Heal JR;  
XX  
DR WPI; 2001-408419/43.  
XX

XX A set of peptide ligands consisting of specific complementary peptides  
CC to proteins encoded by genes of the human genome, useful in an assay  
PT for screening and identifying of one or more novel peptides which are  
PT drug candidates or pro-drugs -  
XX  
PS Example 4; Page 215; 646pp; English.  
XX

XX The invention relates to a set of complementary peptide ligands  
CC generated from the human genome. The complementary peptides  
CC interact with their relevant target proteins encoded in the human  
CC genome. They can be used as reagents in drug discovery and as lead  
CC ligands to facilitate drug design and development. The present  
CC sequence is a complementary peptide provided in the specification.  
XX  
SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;  
Best Local Similarity 55.6%; Pred. No. 79;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLATGSARQ 9  
DB 1 cartgsgrr 9

0;

## RESULT 5

AAG95037  
ID AAG95037 standard; Peptide; 10 AA.  
XX

AC AAG95037;  
XX

DT 18-SEP-2001 (first entry)  
XX

DE Human complementary peptide, SEQ ID NO: 1231.  
XX

KW Human; complementary peptide; ligand; drug discovery; drug design.  
XX

OS Homo sapiens.  
XX

PN WO200142277-A2.  
XX

PD 14-JUN-2001.  
XX

PF 13-DEC-2000; 2000WO-GB04776.  
XX

PR 13-DEC-1999; 99GB-0029464.  
XX

PA (PROT-) PROTEOM LTD.  
XX

PI Roberts GW, Heal JR;  
XX

DR WPI; 2001-408419/43.  
XX

XX A set of peptide ligands consisting of specific complementary peptides  
CC to proteins encoded by genes of the human genome, useful in an assay  
PT for screening and identifying of one or more novel peptides which are  
PT drug candidates or pro-drugs -  
XX

PS Example 4; Page 216; 646pp; English.  
XX

XX The invention relates to a set of complementary peptide ligands  
CC generated from the human genome. The complementary peptides

CC interact with their relevant target proteins encoded in the human  
CC genome. They can be used as reagents in drug discovery and as lead  
CC ligands to facilitate drug design and development. The present  
CC sequence is a complementary peptide provided in the specification.  
XX  
SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;  
Best Local Similarity 55.6%; Pred. No. 79;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLATGSARQ 9  
DB 1 cartgsgrr 9

## RESULT 6

AAG95041  
ID AAG95041 standard; Peptide; 10 AA.  
XX

AC AAG95041;  
XX

DT 18-SEP-2001 (first entry)  
XX

DE Human complementary peptide, SEQ ID NO: 1235.  
XX

KW Human; complementary peptide; ligand; drug discovery; drug design.  
XX

OS Homo sapiens.  
XX

PN WO200142277-A2.  
XX

PD 14-JUN-2001.  
XX

PF 13-DEC-2000; 2000WO-GB04776.  
XX

PR 13-DEC-1999; 99GB-0029464.  
XX

PA (PROT-) PROTEOM LTD.  
XX

PI Roberts GW, Heal JR;  
XX

DR WPI; 2001-408419/43.  
XX

XX A set of peptide ligands consisting of specific complementary peptides  
CC to proteins encoded by genes of the human genome, useful in an assay  
PT for screening and identifying of one or more novel peptides which are  
PT drug candidates or pro-drugs -  
XX

PS Example 4; Page 217; 646pp; English.  
XX

XX The invention relates to a set of complementary peptide ligands  
CC generated from the human genome. The complementary peptides  
CC interact with their relevant target proteins encoded in the human  
CC genome. They can be used as reagents in drug discovery and as lead  
CC ligands to facilitate drug design and development. The present  
CC sequence is a complementary peptide provided in the specification.  
XX  
SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;  
Best Local Similarity 55.6%; Pred. No. 79;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLATGSARQ 9  
DB 1 cartgsgrr 9

## RESULT 7

AAG95045

Thu Jun 6 10:09:31 2002

```

XX PA (PROT-) PROTEOM LTD.
XX XX Roberts GW, Heal JR;
XX PI
XX XX
XX DR WPI; 2001-408419/43.
XX XX
XX PT A set of peptide ligands consisting of specific complementary peptides
XX PT to proteins encoded by genes of the human genome, useful in an assay
XX PT for screening and identifying of one or more novel peptides which are
XX PT drug candidates or pro-drugs.
XX XX
XX PS Example 4; Page 224; 646pp; English.
XX CC The invention relates to a set of complementary peptide ligands
XX CC generated from the human genome. The complementary peptides
XX CC interact with their relevant target proteins encoded in the human
XX CC genome. They can be used as reagents in drug discovery and as lead
XX CC ligands to facilitate drug design and development. The present
XX CC sequence is a complementary peptide provided in the specification.
XX XX
XX SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;
Best Local Similarity 55.6%; Pred. No. 79;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLATGSARQ 9
   | | | | |
Db 2 cartgsgrr 10

RESULT 9
AAG95093
ID AAG95093 standard; Peptide; 10 AA.
XX AC AAG95093;
XX AC
XX DT 18-SEP-2001 (first entry)
XX XX
XX DE Human complementary peptide; ligand; drug discovery; drug design.
XX DE
XX KW Human; complementary peptide; ligand; drug discovery; drug design.
XX KW
XX OS Homo sapiens.
XX OS
XX PN WO200142277-A2.
XX PN
XX PD 14-JUN-2001.
XX PD
XX PF 13-DEC-2000; 2000WO-GB04776.
XX PF
XX PR 13-DEC-1999; 99GB-0029464.
XX PR
XX PA (PROT-) PROTEOM LTD.
XX PA
XX PI Roberts GW, Heal JR;
XX PI
XX XX WPI; 2001-408419/43.
XX XX
XX PT A set of peptide ligands consisting of specific complementary peptides
XX PT to proteins encoded by genes of the human genome, useful in an assay
XX PT for screening and identifying of one or more novel peptides which are
XX PT drug candidates or pro-drugs.
XX XX
XX PS Example 4; Page 217; 646pp; English.
XX CC The invention relates to a set of complementary peptide ligands
XX CC generated from the human genome. The complementary peptides
XX CC interact with their relevant target proteins encoded in the human
XX CC genome. They can be used as reagents in drug discovery and as lead
XX CC ligands to facilitate drug design and development. The present
XX CC sequence is a complementary peptide provided in the specification.
XX XX
XX SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;
Best Local Similarity 55.6%; Pred. No. 79;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLATGSARQ 9
   | | | | |
Db 1 cartgsgrr 9

RESULT 8
AAG95091
ID AAG95091 standard; Peptide; 10 AA.
XX AC AAG95091;
XX AC
XX DT 18-SEP-2001 (first entry)
XX XX
XX DE Human complementary peptide; ligand; drug discovery; drug design.
XX DE
XX KW Human; complementary peptide; ligand; drug discovery; drug design.
XX KW
XX OS Homo sapiens.
XX OS
XX PN WO200142277-A2.
XX PN
XX PD 14-JUN-2001.
XX PD
XX PF 13-DEC-2000; 2000WO-GB04776.
XX PF
XX PR 13-DEC-1999; 99GB-0029464.
XX PR

```

CC sequence is a complementary peptide provided in the specification.  
 XX  
 SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;  
 Best Local Similarity 55.6%; Pred. No. 79;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLATGSARQ 9  
 | | | | |  
 Db 2 cartgsgrr 10

## RESULT 10

AAG95095  
 ID AAG95095 standard; Peptide; 10 AA.

AC AAG95095;

XX 18-SEP-2001 (first entry)

XX Human complementary peptide, SEQ ID NO: 1289.

DE Human; complementary peptide; ligand; drug discovery; drug design.

KW Homo sapiens.

OS WO200142277-A2.

XX 14-JUN-2001.

PF 13-DEC-2000; 2000WO-GB04776.

PR 13-DEC-1999; 99GB-0029464.

XX (PROT-) PROTEOM LTD.

PI Roberts GW, Heal JR;

XX WPI; 2001-408419/43.

XX A set of peptide ligands consisting of specific complementary peptides  
 PT to proteins encoded by genes of the human genome, useful in an assay  
 PT for screening and identifying of one or more novel peptides which are  
 PT drug candidates or pro-drugs -

XX Example 4; Page 225; 646pp; English.

XX The invention relates to a set of complementary peptide ligands  
 CC generated from the human genome. The complementary peptides  
 CC interact with their relevant target proteins encoded in the human  
 CC genome. They can be used as reagents in drug discovery and as lead  
 CC ligands to facilitate drug design and development. The present  
 CC sequence is a complementary peptide provided in the specification.

SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;  
 Best Local Similarity 55.6%; Pred. No. 79;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLATGSARQ 9  
 | | | | |  
 Db 2 cartgsgrr 10

## RESULT 11

AAG95097  
 ID AAG95097 standard; Peptide; 10 AA.

XX AAG95097;

XX 18-SEP-2001 (first entry)  
 XX Human complementary peptide, SEQ ID NO: 1291.  
 DE Human; complementary peptide; ligand; drug discovery; drug design.  
 KW Homo sapiens.  
 XX WO200142277-A2.  
 XX 14-JUN-2001.  
 XX 13-DEC-2000; 2000WO-GB04776.  
 XX 13-DEC-1999; 99GB-0029464.  
 XX (PROT-) PROTEOM LTD.  
 XX Roberts GW, Heal JR;  
 XX WPI; 2001-408419/43.

XX A set of peptide ligands consisting of specific complementary peptides  
 PT to proteins encoded by genes of the human genome, useful in an assay  
 PT for screening and identifying of one or more novel peptides which are  
 PT drug candidates or pro-drugs -

XX Example 4; Page 225; 646pp; English.

XX The invention relates to a set of complementary peptide ligands  
 CC generated from the human genome. The complementary peptides  
 CC interact with their relevant target proteins encoded in the human  
 CC genome. They can be used as reagents in drug discovery and as lead  
 CC ligands to facilitate drug design and development. The present  
 CC sequence is a complementary peptide provided in the specification.

SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;  
 Best Local Similarity 55.6%; Pred. No. 79;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLATGSARQ 9  
 | | | | |  
 Db 1 cartgsgrr 9

## RESULT 12

AAG95099  
 ID AAG95099 standard; Peptide; 10 AA.

XX AAG95099;

XX 18-SEP-2001 (first entry)

XX Human complementary peptide, SEQ ID NO: 1293.

XX Human; complementary peptide; ligand; drug discovery; drug design.  
 XX Homo sapiens.

XX WO200142277-A2.

XX 14-JUN-2001.

XX 13-DEC-2000; 2000WO-GB04776.

XX 13-DEC-1999; 99GB-0029464.

XX (PROT-) PROTEOM LTD.

Thu Jun 6 10:09:31 2002

us-08-881-509-8.closed.rag

PI Roberts GW, Heal JR;  
XX WPI; 2001-408419/43.

XX A set of peptide ligands consisting of specific complementary peptides  
XX to proteins encoded by genes of the human genome, useful in an assay  
PT for screening and identifying of one or more novel peptides which are  
PT drug candidates or pro-drugs -

XX Example 4; Page 225; 646pp; English.

XX The invention relates to a set of complementary peptide ligands  
CC generated from the human genome. The complementary peptides  
CC interact with their relevant target proteins encoded in the human  
CC genome. They can be used as reagents in drug discovery and as lead  
CC ligands to facilitate drug design and development. The present  
CC sequence is a complementary peptide provided in the specification.

XX Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;  
Best Local Similarity 55.6%; Pred. No. 79;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 CLATGSARQ 9  
| | | | |  
Db 1 cartgsgrr 9

RESULT 13

AAG95101  
ID AAG95101 standard; Peptide; 10 AA.

XX AAG95101;

XX 18-SEP-2001 (first entry)

XX Human complementary peptide, SEQ ID NO: 1295.

XX Human; complementary peptide; ligand; drug discovery; drug design.

XX Homo sapiens.

XX WO200142277-A2.

XX 14-JUN-2001.

XX 13-DEC-2000; 2000WO-GB04776.

XX 13-DEC-1999; 99GB-0029464.

XX (PROT-) PROTEOM LTD.

XX Roberts GW, Heal JR;

XX WPI; 2001-408419/43.

XX A set of peptide ligands consisting of specific complementary peptides  
PT to proteins encoded by genes of the human genome, useful in an assay  
PT for screening and identifying of one or more novel peptides which are  
PT drug candidates or pro-drugs -

XX Example 4; Page 226; 646pp; English.

XX The invention relates to a set of complementary peptide ligands  
CC generated from the human genome. The complementary peptides  
CC interact with their relevant target proteins encoded in the human  
CC genome. They can be used as reagents in drug discovery and as lead  
CC ligands to facilitate drug design and development. The present  
CC sequence is a complementary peptide provided in the specification.

XX Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;  
Best Local Similarity 55.8%; Pred. No. 79;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 CLATGSARQ 9  
| | | | |  
Db 1 cartgsgrr 9

RESULT 14

AAG95115  
ID AAG95115 standard; Peptide; 10 AA.

XX AAG95115;

XX 18-SEP-2001 (first entry)

XX Human complementary peptide, SEQ ID NO: 1309.

XX Human; complementary peptide; ligand; drug discovery; drug design.

XX Homo sapiens.

XX WO200142277-A2.

XX 14-JUN-2001.

XX 13-DEC-2000; 2000WO-GB04776.

XX 13-DEC-1999; 99GB-0029464.

XX (PROT-) PROTEOM LTD.

XX Roberts GW, Heal JR;

XX WPI; 2001-408419/43.

XX A set of peptide ligands consisting of specific complementary peptides  
PT to proteins encoded by genes of the human genome, useful in an assay  
PT for screening and identifying of one or more novel peptides which are  
PT drug candidates or pro-drugs -

XX Example 4; Page 228; 646pp; English.

XX The invention relates to a set of complementary peptide ligands  
CC generated from the human genome. The complementary peptides  
CC interact with their relevant target proteins encoded in the human  
CC genome. They can be used as reagents in drug discovery and as lead  
CC ligands to facilitate drug design and development. The present  
CC sequence is a complementary peptide provided in the specification.

XX Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;  
Best Local Similarity 55.6%; Pred. No. 79;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 CLATGSARQ 9  
| | | | |  
Db 1 cartgsgrr 9

RESULT 15

AAG95141  
ID AAG95141 standard; Peptide; 10 AA.

XX AAG95141;

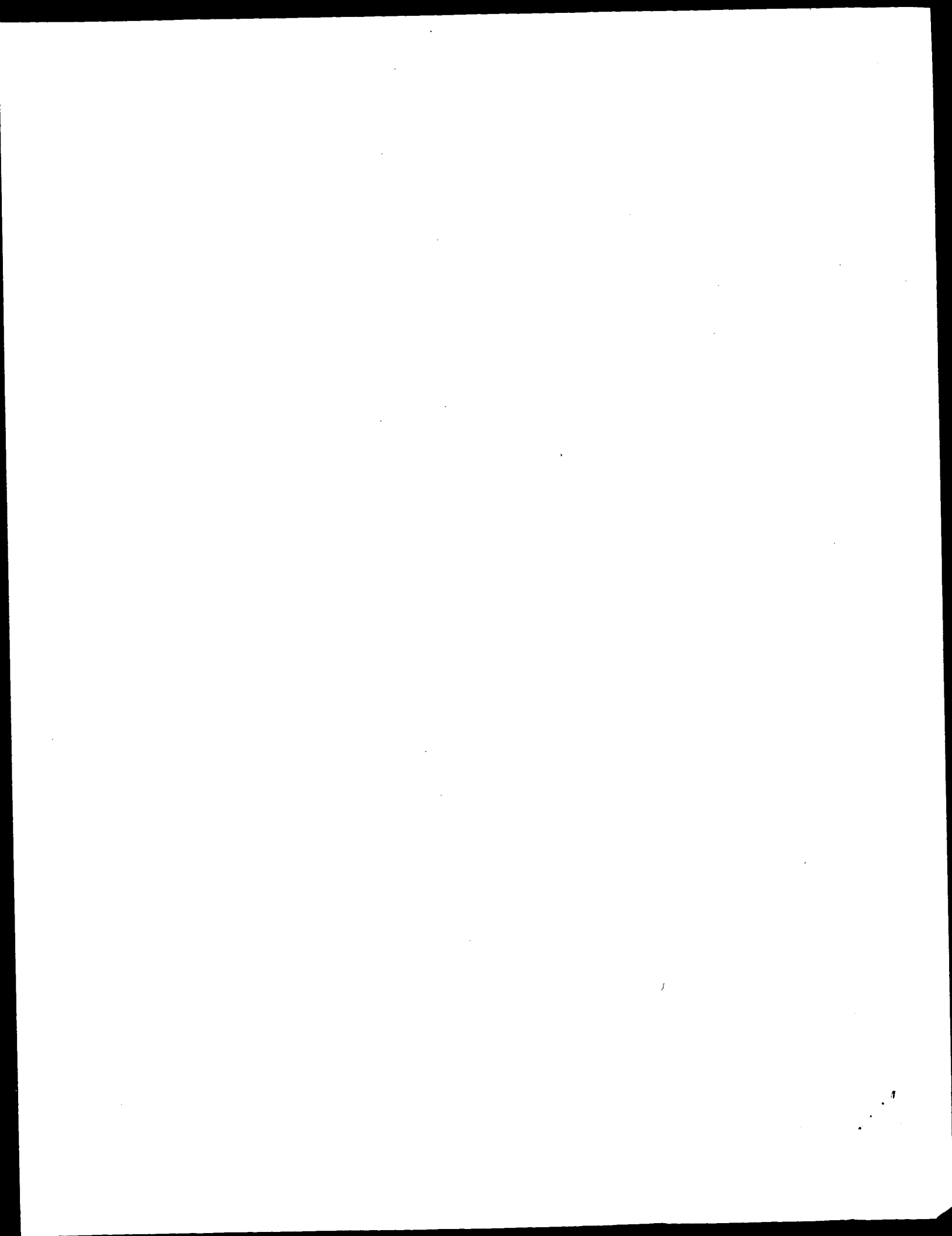
XX 18-SEP-2001 (first entry)

XX

DE Human complementary peptide, SEQ ID NO: 1335.  
XX  
KW Human; complementary peptide; ligand; drug discovery; drug design.  
XX  
OS Homo sapiens.  
XX  
PN WO200142277-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 13-DEC-2000; 2000WO-GB04776.  
XX  
PR 13-DEC-1999; 99GB-0029464.  
XX  
PA (PROT-) PROTEOM LTD.  
XX  
PI Roberts GW, Heal JR;  
XX  
DR WPI; 2001-408419/43.  
XX  
PT A set of peptide ligands consisting of specific complementary peptides  
PT to proteins encoded by genes of the human genome, useful in an assay  
PT for screening and identifying of one or more novel peptides which are  
PT drug candidates or pro-drugs -  
XX  
PS Example 4; Page 232; 646pp; English.  
XX  
CC The invention relates to a set of complementary peptide ligands  
CC generated from the human genome. The complementary peptides  
CC interact with their relevant target proteins encoded in the human  
CC genome. They can be used as reagents in drug discovery and as lead  
CC ligands to facilitate drug design and development. The present  
CC sequence is a complementary peptide provided in the specification.  
XX  
SQ Sequence 10 AA;

Query Match 45.9%; Score 28; DB 22; Length 10;  
Best Local Similarity 55.6%; Pred. No. 79;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CLATGSARQ 9  
| | | | |  
Db 1 cartsgrr 9

Search completed: June 6, 2002, 06:27:11  
Job time: 204 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 06:25:32 ; Search time 12.84 Seconds  
(without alignments)  
22.828 Million cell updates/sec

Title: US-08-881-509-8  
Perfect score: 61  
Sequence: 1 CLATGSARQLTF 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 83765

Minimum DB seq length: 0  
Maximum DB seq length: 12

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/6C.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/6D.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	45.9	12	6	5424218-10
2	27	44.3	8	2	US-08-751-767A-80
3	26	42.6	12	2	US-08-950-866-8
4	25	41.0	9	1	US-08-049-794-27
5	25	41.0	9	1	US-08-496-847-27
6	25	41.0	9	2	US-08-742-774-27
7	25	41.0	9	2	US-08-675-354-27
8	25	41.0	9	2	US-08-965-918-27
9	25	41.0	9	2	US-09-138-439-27
10	25	41.0	9	3	US-08-613-400A-27
11	25	41.0	9	3	US-09-298-017-27
12	25	41.0	9	4	US-08-312-979A-27
13	25	41.0	10	1	US-08-214-650-33
14	24	39.3	7	6	5378464-26
15	24	39.3	10	3	US-08-159-339A-720
16	24	39.3	12	1	US-07-834-902-6
17	24	39.3	12	1	US-08-018-904-6
18	24	39.3	12	1	US-08-294-675A-6
19	24	39.3	12	1	US-08-397-633A-95
20	24	39.3	12	2	US-08-950-866-3
21	23	37.7	9	2	US-08-340-283-134
22	22	36.1	9	6	5217869-105
23	22	36.1	10	1	US-08-250-789A-107
24	22	36.1	10	1	US-08-348-333-57
25	22	36.1	10	4	US-09-230-548-6
26	22	36.1	10	4	US-09-020-065A-26
27	22	36.1	10	4	US-09-020-065A-26

28	22	36.1	10	5	PCT-US91-07506-57	Sequence 57, Appl
29	22	36.1	12	1	US-08-548-540-151	Sequence 151, App
30	22	36.1	12	1	US-08-419-903A-6	Sequence 6, Appl
31	22	36.1	12	2	US-08-292-968-2	Sequence 2, Appl
32	22	36.1	12	2	US-08-467-974-2	Sequence 2, Appl
33	22	36.1	12	3	US-08-467-536-2	Sequence 2, Appl
34	22	36.1	12	3	US-08-467-976-2	Sequence 2, Appl
35	22	36.1	12	4	US-09-082-514-2	Sequence 2, Appl
36	22	36.1	12	5	PCT-US96-09809-151	Sequence 151, App
37	21	34.4	7	4	US-09-333-729A-18	Sequence 18, Appl
38	21	34.4	8	2	US-08-669-284B-32	Sequence 32, Appl
39	21	34.4	9	2	US-08-340-283-30	Sequence 30, Appl
40	21	34.4	9	2	US-08-340-283-104	Sequence 104, App
41	21	34.4	9	2	US-08-340-283-127	Sequence 127, App
42	21	34.4	9	4	US-09-258-754-333	Sequence 333, App
43	21	34.4	9	4	US-09-042-107-333	Sequence 333, App
44	21	34.4	9	4	US-08-584-008A-5	Sequence 5, Appl
45	21	34.4	9	4	US-09-492-543-33	Sequence 33, Appl

## ALIGNMENTS

RESULT 1  
5424218-10  
; Patent No. 5424218  
; APPLICANT: MILJANICH, GEORGE P.; BITNER, ROBERT S.; BOWERSOX, STEPHEN S.; FOX, JAMES A.; VALENTINO, KAREN L.; YAMASHIRO, DONALD H.  
; TITLE OF INVENTION: SCREENING METHOD FOR NEUROPROTECTIVE COMPOUNDS  
; NUMBER OF SEQUENCES: 21  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/147,714  
; FILING DATE: 04-NOV-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 855,269  
; FILING DATE: 23-MAR-1992  
; APPLICATION NUMBER: 561,766  
; FILING DATE: 02-AUG-1990  
; APPLICATION NUMBER: 440,094  
; FILING DATE: 22-NOV-1989  
; SEQ ID NO: 10:  
; LENGTH: 12  
5424218-10

Query Match 45.9% Score 28; DB 6; Length 12;  
Best Local Similarity 62.5%; Pred. No. 19;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 CLATGSAR 8  
Db 1 CCCTGSCR 8

RESULT 2  
US-08-751-767A-80  
; Sequence 80, Application US/08751767A  
; Patent No. 5994104  
; GENERAL INFORMATION:  
; APPLICANT: ANDERSON, ROBERT J.  
; APPLICANT: GRANT, HUGH  
; APPLICANT: MACDONALD, IAN D.  
; TITLE OF INVENTION: INTERLUKIN-12 FUSION PROTEIN  
; NUMBER OF SEQUENCES: 80  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHVE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22201  
; COMPUTER READABLE FORM: Floppy disk  
; MEDIUM TYPE: Floppy disk

us-08-881-509-8.closed.ra1

Thu Jun 6 10:09:31 2002

Query Match 42.6%; Score 26; DB 2; Length 12;  
Best Local Similarity 62.5%; Pred. No. 47;  
Matches 5; Conservative 1; Mismatches 0; Gaps 0;

QY 1 CLATGSAR 8  
| | | | |  
Db 1 CRDTGSSR 8

RESULT 4  
US-08-049-794-27  
; Sequence 27, Application US/08049794  
; Patent No. 5587454  
; GENERAL INFORMATION:  
; APPLICANT: JUSTICE, ALAN  
; APPLICANT: SINGH, TEJINDER  
; APPLICANT: GOHIL, KISHOR C  
; APPLICANT: VALENTINO, KAREN L  
; APPLICANT: MILJANICH, GEORGE P  
; TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND  
; TITLE OF INVENTION: ENHANCING OPIATE ANALGESIA  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Law Offices of Peter Dehlinger  
; STREET: 350 Cambridge Avenue, Suite 300  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/049,794  
; FILING DATE: 19930415  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/814,759  
; FILING DATE: 30-DEC-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stratford, Carol A.  
; REGISTRATION NUMBER: 34,444  
; REFERENCE/DOCKET NUMBER: 5865-0009.30  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 2 FRAGMENT, PAGE  
; INDIVIDUAL ISOLATE: 32  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 4  
; OTHER INFORMATION: /note= "where X is hydroxyproline"  
US-08-049-794-27

Query Match 41.0%; Score 25; DB 1; Length 9;  
Best Local Similarity 57.1%; Pred. No. 1.7e+05;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLATGSA 7  
| | | | |  
Db 1 CLSXGSS 7

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/751,767A  
FILING DATE: 08-NOV-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: SADOFF, B.J.  
REGISTRATION NUMBER: 36,663  
REFERENCE/DOCKET NUMBER: 117-221  
TELEPHONE: 7038164091  
TELEFAX: 7038164100  
INFORMATION FOR SEQ ID NO: 80:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-751-767A-80

Query Match 44.3%; Score 27; DB 2; Length 8;  
Best Local Similarity 62.5%; Pred. No. 1.7e+05;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CLATGSAR 8  
| | | | |  
Db 1 CAATGRVR 8

RESULT 3  
US-08-950-866-8  
; Sequence 8, Application US/08950866  
; Patent No. 5840296  
; GENERAL INFORMATION:  
; APPLICANT: Raines, Ronald T  
; TITLE OF INVENTION: Engineered Cytotoxic Ribonuclease A  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/950,866  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seav, Nicholas J.  
; REGISTRATION NUMBER: 27386  
; REFERENCE/DOCKET NUMBER: 960296.94592  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-950-866-8



RESULT 5  
US-08-496-847-27  
; Sequence 27, Application US/08496847  
; Patent No. 5795864  
; GENERAL INFORMATION:  
; APPLICANT: Amstutz, Gary A.  
; APPLICANT: Bowersox, Stephen S.  
; APPLICANT: Gohil, Kishorchandra  
; APPLICANT: Adriaenssens, Peter I.  
; APPLICANT: Kristipati, Ramasharma  
; TITLE OF INVENTION: METHODS AND  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94306-1546  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/496.847  
; FILING DATE: 27-JUN-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stratford, Carol A.  
; REGISTRATION NUMBER: 34,444  
; REFERENCE/DOCKET NUMBER: 5865-0009.31  
; TELEPHONE: 650-324-0880  
; TELEFAX: 650-324-0960  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 2 FRAGMENT  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 4  
; OTHER INFORMATION: /note= "where X is hydroxyproline"  
US-08-496-847-27

Query Match 41.0%; Score 25; DB 1; Length 9;  
Best Local Similarity 57.1%; Pred. No. 1.7e+05;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 CLATGSA 7  
Db 1 CLSXGSS 7

RESULT 6  
US-08-742-774-27  
; Sequence 27, Application US/08742774  
; Patent No. 5824645  
; GENERAL INFORMATION:  
; APPLICANT: JUSTICE, ALAN  
; APPLICANT: SINGH, TEJINDER  
; APPLICANT: GOHIL, KISHOR C  
; APPLICANT: VALENTINO, KAREN L  
; APPLICANT: MILJANICH, GEORGE P  
; TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND

; TITLE OF INVENTION: ENHANCING OPIATE ANALGESIA  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Law Offices of Peter Dehlinger  
; STREET: 350 Cambridge Avenue, Suite 300  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/742.774  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/675.354  
; FILING DATE: 03-JUL-1996  
; APPLICATION NUMBER: US/08/049.794  
; FILING DATE: 1993-APR-15  
; APPLICATION NUMBER: US 07/814.759  
; FILING DATE: 30-DEC-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stratford, Carol A.  
; REGISTRATION NUMBER: 34,444  
; REFERENCE/DOCKET NUMBER: 5865-0009.30  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 2 FRAGMENT, PAGE  
; INDIVIDUAL ISOLATE: 32  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 4  
; OTHER INFORMATION: /note= "where X is hydroxyproline"  
US-08-742-774-27

Query Match 41.0%; Score 25; DB 2; Length 9;  
Best Local Similarity 57.1%; Pred. No. 1.7e+05;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 CLATGSA 7  
Db 1 CLSXGSS 7

RESULT 7  
US-08-675-354-27  
; Sequence 27, Application US/08675354  
; Patent No. 5859186  
; GENERAL INFORMATION:  
; APPLICANT: JUSTICE, ALAN  
; APPLICANT: SINGH, TEJINDER  
; APPLICANT: GOHIL, KISHOR C  
; APPLICANT: VALENTINO, KAREN L  
; APPLICANT: MILJANICH, GEORGE P  
; TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Law Offices of Peter Dehlinger

Thu Jun 6 10:09:31 2002

us-08-881-509-8.closed.ra1

```
STREET: 350 Cambridge Avenue, Suite 300
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/675.354
APPLICATION NUMBER: US/08/675.354
FILING DATE: 03-JUL-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/049.794
FILING DATE: 1993-APR-15
APPLICATION NUMBER: US 07/814,759
FILING DATE: 30-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Stratford, Carol A.
REGISTRATION NUMBER: 34,444
REFERENCE/DOCKET NUMBER: 5865-0009.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 2 FRAGMENT, PAGE
INDIVIDUAL ISOLATE: 32
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note= "where X is hydroxyproline"
US-08-675-354-27

Query Match 41.0%; Score 25; DB 2; Length 9;
Best Local Similarity 57.1%; Pred. No. 1.7e+05;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLATGSA 7
   ||: ||:
Db 1 CLSXGSS 7

RESULT 8
US-08-965-918-27
; Sequence 27, Application US/08965918
; Patent No. 5891849
; GENERAL INFORMATION:
; APPLICANT: Amstutz, Gary A.
; APPLICANT: Bowersox, Stephen S.
; APPLICANT: Gohil, Kishorchandra
; APPLICANT: Adriaenssens, Peter I.
; APPLICANT: Kristipati, Ramasharma
; TITLE OF INVENTION: METHODS AND FORMULATIONS FOR PREVENTING
; PROGRESSION OF NEUROPATHIC PAIN
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94306-1546
; COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA: US/08/965,918
APPLICATION NUMBER: US/08/965,918
FILING DATE: 07-NOV-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Mohr, Judy M.
REGISTRATION NUMBER: 38,563
REFERENCE/DOCKET NUMBER: 5865-0009.34
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-0880
TELEFAX: 650-324-0960
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 2 FRAGMENT
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note= "where X is hydroxyproline"
US-08-965-918-27

Query Match 41.0%; Score 25; DB 2; Length 9;
Best Local Similarity 57.1%; Pred. No. 1.7e+05;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLATGSA 7
   ||: ||:
Db 1 CLSXGSS 7

RESULT 9
US-09-138-439-27
; Sequence 27, Application US/09138439
; Patent No. 5994305
; GENERAL INFORMATION:
; APPLICANT: JUSTICE, ALAN
; APPLICANT: SINGH, TEJINDER
; APPLICANT: GOHIL, KISHOR C.
; APPLICANT: VALENTINO, KAREN L.
; APPLICANT: MILJANICH, GEORGE P.
; TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND
; ENHANCING OPIATE ANALGESIA
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Peter Dehlinger
; STREET: 350 Cambridge Avenue, Suite 300
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/138,439
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/049,794
; FILING DATE: 1993-04-15
; APPLICATION NUMBER: US 07/814,759
```

;; FILING DATE: 30-DEC-1991  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Stratford, Carol A.  
;; REGISTRATION NUMBER: 34,444  
;; REFERENCE/DOCKET NUMBER: 5865-0009.30  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 324-0880  
;; TELEFAX: (415) 324-0960  
;; INFORMATION FOR SEQ ID NO: 27:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 9 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; HYPOTHETICAL: NO  
;; ORIGINAL SOURCE:  
;; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 2 FRAGMENT, PAGE  
;; INDIVIDUAL ISOLATE: 32  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 4  
;; OTHER INFORMATION: /note= "where x is hydroxyproline"  
;; US-09-138-439-27

Query Match 41.0%; Score 25; DB 2; Length 9;  
Best Local Similarity 57.1%; Pred. No. 1.7e+05;  
Matches 4; Conservative 2; Mismatches 1; Indels 0;

QY 1 CLATGSA 7  
DB 1 CLSXGSS 7

RESULT 10  
US-08-613-400A-27  
; Sequence 27, Application US/08613400A  
; Patent No. 6054429  
; GENERAL INFORMATION:  
; APPLICANT: Bowersox, S. Scott  
; APPLICANT: Gadbois, Theresa  
; APPLICANT: Pettus, Mark, R.  
; APPLICANT: Luther, Robert, R.  
; TITLE OF INVENTION: IMPROVED EPIDURAL  
; TITLE OF INVENTION: METHOD OF PRODUCING ANALGESIA  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94306-1546  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/613,400A  
; FILING DATE: 08-MAR-1996  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stratford, Carol A.  
; REGISTRATION NUMBER: 34,444  
; REFERENCE/DOCKET NUMBER: 5865-0019  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-324-0960  
; TELEFAX: 650-324-0960  
; INFORMATION FOR SEQ ID NO: 27:

;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 9 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; HYPOTHETICAL: NO  
;; ORIGINAL SOURCE:  
;; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 2 FRAGMENT,  
;; INDIVIDUAL ISOLATE: page 33, line29 to page 34, line4  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 4  
;; OTHER INFORMATION: /note= "where x is hydroxyproline"  
;; US-08-613-400A-27

Query Match 41.0%; Score 25; DB 3; Length 9;  
Best Local Similarity 57.1%; Pred. No. 1.7e+05;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLATGSA 7  
DB 1 CLSXGSS 7

RESULT 11  
US-09-298-017-27  
; Sequence 27, Application US/09298017  
; Patent No. 6087091  
; GENERAL INFORMATION:  
; APPLICANT: JUSTICE, ALAN  
; APPLICANT: SINGH, TEJINDER  
; APPLICANT: GOHIL, KISHOR C  
; APPLICANT: VALENTINO, KAREN L  
; APPLICANT: MILJANICH, GEORGE P  
; TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND  
; TITLE OF INVENTION: ENHANCING OPIATE ANALGESIA  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Law Offices of Peter Dehlinger  
; STREET: 350 Cambridge Avenue, Suite 300  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/298,017  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/049,794  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stratford, Carol A.  
; REGISTRATION NUMBER: 34,444  
; REFERENCE/DOCKET NUMBER: 5865-0009.30  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 2 FRAGMENT, PAGE

```

; INDIVIDUAL ISOLATE: 32
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 4
; OTHER INFORMATION: /note= "where X is hydroxyproline"
US-09-298-017-27

Query Match      41.0%; Score 25; DB 3; Length 9;
Best Local Similarity 57.1%; Pred. No. 1.7e+05;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLATGSA 7
Db 1 CLSXGSS 7

RESULT 12
US-09-392-979A-27
; Sequence 27, Application US/09392979A
; Patent No. 6136786
; GENERAL INFORMATION:
; APPLICANT: JUSTICE, ALAN
; APPLICANT: SINGH, TEJINDER
; APPLICANT: GOHIL, KISHOR C.
; APPLICANT: VALENTINO, KAREN L.
; APPLICANT: MILJANICH, GEORGE P.
; TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND
; TITLE OF INVENTION: ENHANCING OPIATE ANALGESIA
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Peter Dehlinger
; STREET: 350 Cambridge Avenue, Suite 300
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/392,979A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/049,794
; FILING DATE: 1993-04-15
; APPLICATION NUMBER: US 07/814,759
; FILING DATE: 30-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Stratford, Carol A.
; REGISTRATION NUMBER: 34,444
; REFERENCE/DOCKET NUMBER: 5865-0009.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 2 FRAGMENT, PAGE
; INDIVIDUAL ISOLATE: 32
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 4
; OTHER INFORMATION: /note= "where X is hydroxyproline"
US-09-392-979A-27

Query Match      41.0%; Score 25; DB 3; Length 9;
Best Local Similarity 57.1%; Pred. No. 1.7e+05;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLATGSA 7
Db 1 CLSXGSS 7

RESULT 13
US-08-214-650-33
; Sequence 33, Application US/08214650
; Patent No. 5709995
; GENERAL INFORMATION:
; APPLICANT: Chisari, Francis V.
; APPLICANT: Cerny, Andreas
; TITLE OF INVENTION: PEPTIDES FOR INDUCING CYTOTOXIC T
; TITLE OF INVENTION: LYMPHOCYTE RESPONSES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voit & Mayer
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,650
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Silver, Donald J.
; REGISTRATION NUMBER: 37552
; REFERENCE/DOCKET NUMBER: 61230
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5600
; TELEFAX: (312) 616-5700
; TELEX: 25-3533
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-214-650-33

Query Match      41.0%; Score 25; DB 1; Length 10;
Best Local Similarity 80.0%; Pred. No. 60;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLATG 5
Db 1 CLSTG 5

RESULT 14
US-08-214-650-33
; Patent No. 5378464
; APPLICANT: MCEVER, RODGER P.
; TITLE OF INVENTION: MODULATION OF INFLAMMATORY RESPONSES
; BY ADMINISTRATION OF GMP-140 OR ANTIBODY TO GMP-140
; NUMBER OF SEQUENCES: 32
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/320,408
US-08-214-650-33

Query Match      41.0%; Score 25; DB 1; Length 10;
Best Local Similarity 80.0%; Pred. No. 60;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLATG 5
Db 1 CLSTG 5

RESULT 14
US-08-214-650-33
; Patent No. 5378464
; APPLICANT: MCEVER, RODGER P.
; TITLE OF INVENTION: MODULATION OF INFLAMMATORY RESPONSES
; BY ADMINISTRATION OF GMP-140 OR ANTIBODY TO GMP-140
; NUMBER OF SEQUENCES: 32
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/320,408
US-08-214-650-33
```

; FILING DATE: 08-MAR-1989  
; SEQ ID NO: 26;  
; LENGTH: 7  
5378464-26

Query Match 39.3%; Score 24; DB 6; Length 7;  
Best Local Similarity 80.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLATG 5  
Db 1 CLASG 5

RESULT 15  
US-08-159-339A-720  
; Sequence 720, Application US/08159339A  
; Patent No. 6037135  
; GENERAL INFORMATION:  
; APPLICANT: Kubo, Ralph T.  
; APPLICANT: Grey, Howard M.  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Celis, Esteban  
; TITLE OF INVENTION: HLA Binding peptides and Their  
; NUMBER OF SEQUENCES: 1254  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/159,339A  
; FILING DATE: 29-NOV-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/926,666  
; FILING DATE: 07-AUG-1992  
; APPLICATION NUMBER: US 08/027,746  
; FILING DATE: 05-MAR-1993  
; APPLICATION NUMBER: US 08/103,396  
; FILING DATE: 06-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weber, Ellen Lauver  
; REGISTRATION NUMBER: 32,762  
; REFERENCE/DOCKET NUMBER: 018623-005030US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 720:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-159-339A-720

Query Match 39.3%; Score 24; DB 3; Length 10;  
Best Local Similarity 57.1%; Pred. No. 95;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 TGSARQL 10

Db 4 TGAVRQI 10

Search completed: June 6, 2002, 06:27:31  
Job time: 119 sec

us-08-881-509-8.closed.ra1

Thu Jun 6 10:09:31 2002